

Washington Law Review

Volume 68 | Number 3

7-1-1993

Statistical Proof of Discrimination: Beyond "Damned Lies"

Kingsley R. Browne

Follow this and additional works at: <https://digitalcommons.law.uw.edu/wlr>



Part of the [Labor and Employment Law Commons](#)

Recommended Citation

Kingsley R. Browne, *Statistical Proof of Discrimination: Beyond "Damned Lies"*, 68 Wash. L. Rev. 477 (1993).

Available at: <https://digitalcommons.law.uw.edu/wlr/vol68/iss3/2>

This Article is brought to you for free and open access by the Law Reviews and Journals at UW Law Digital Commons. It has been accepted for inclusion in Washington Law Review by an authorized editor of UW Law Digital Commons. For more information, please contact cnyberg@uw.edu.

STATISTICAL PROOF OF DISCRIMINATION: BEYOND "DAMNED LIES"

Kingsley R. Browne*

Abstract: Evidence that an employer's work force contains fewer minorities or women than would be expected if selection were random with respect to race and sex has been taken as powerful—and often sufficient—evidence of systematic intentional discrimination. In relying on this kind of statistical evidence, courts have made two fundamental errors. The first error is assuming that statistical analysis can reveal the probability that observed work-force disparities were produced by chance. This error leads courts to exclude chance as a cause when such a conclusion is unwarranted. The second error is assuming that, except for random deviations, the work force of a nondiscriminating employer would mirror the racial and sexual composition of the relevant labor force. This assumption has led courts inappropriately to shift the burden of proof to employers in pattern-or-practice cases once a statistical disparity is shown. Recognition of these two errors suggests that the role of statistical evidence in discrimination cases should be greatly reduced.

I. INTRODUCTION

It has been said¹—and frequently repeated²—that “[i]n the problem of . . . discrimination, statistics often tell much, and Courts listen.”³ As in the parlor game “Gossip,” however, the message ultimately received is usually not the message transmitted. The method of proving discrimination through statistical proof is based upon faulty statistical and factual assumptions, and because of misconceived interpretations of the meaning of statistical evidence, courts have developed evidentiary doctrines that have the effect of improperly shifting the burden of proof to defendants in discrimination cases. It is the thesis of this Article that statistical evidence of intentional discrimination should be abandoned as a primary method of proof and should become, at most, merely an adjunct to evidence that specific persons have been subjected to discrimination.

As Laurence Tribe has observed, “the costs of abusing a technique must be reckoned among the costs of using it at all to the extent that

* Associate Professor of Law, Wayne State University Law School; B.A., George Washington University (1975); M.A., University of Colorado (1976); J.D., University of Denver (1982). The author would like to thank Joseph Grano, Michael McIntyre, Stephen Schulman, and Robert Sedler for their helpful comments on earlier drafts of this Article.

1. *Alabama v. United States*, 304 F.2d 583, 586 (5th Cir.), *aff'd*, 371 U.S. 37 (1962).

2. *See, e.g., Lams v. General Waterworks Corp.*, 766 F.2d 386, 391 (8th Cir. 1985); *Bryant v. International Sch. Servs., Inc.*, 675 F.2d 562, 573 (3d Cir. 1982); *Blake v. City of Los Angeles*, 595 F.2d 1367, 1375 n.4 (9th Cir. 1979), *cert. denied*, 446 U.S. 928 (1980).

3. *Alabama v. United States*, 304 F.2d at 586.

the latter creates risks of the former.”⁴ The costs of using statistics to prove discrimination are significant because statistics are systematically misinterpreted and their meaning exaggerated; the associated costs are all the more disturbing, since the utility of statistics, if used properly in such cases, is insignificant. One might expect that as the shortcomings of statistical evidence became more obvious enthusiasm for such evidence would wane. Unfortunately, that has not been the case. Instead, limitations on the utility of statistical evidence have been relied upon not to limit the use of such evidence, but rather to modify evidentiary rules in ways that tend to obscure its lack of usefulness, and in the process to shift the burden of proof, albeit surreptitiously, to defendants in employment discrimination cases.

There are two basic theories of discrimination under Title VII—the “disparate impact” theory and the “disparate treatment” theory—both of which may involve statistical proof.⁵ Under disparate-impact theory, the plaintiff challenges a facially neutral employment practice on the ground that it produces an adverse—if only inadvertent—effect on a protected group.⁶ Under disparate-treatment theory, the plaintiff alleges that the employer has intentionally treated one or more members of a protected category differently from members of other groups. A disparate-treatment claim may be brought as an individual action—where the plaintiff complains about his own specific treatment⁷—or it may be brought as a class action or pattern-or-practice case⁸—in which case the plaintiffs (either individuals or the Equal Employment

4. Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1331 (1971).

5. These modes of analysis have also been used in claims brought under the Age Discrimination in Employment Act (ADEA), although the Supreme Court has never endorsed the use of disparate-impact theory under the ADEA. See *Markham v. Geller*, 451 U.S. 945, 948 (1981) (Rehnquist, J., dissenting) (denying certiorari); *Metz v. Transit Mix, Inc.*, 828 F.2d 1202, 1220 (7th Cir. 1987) (Easterbrook, J., dissenting).

6. See, e.g., *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971).

7. See, e.g., *McDonnell Douglas Corp. v. Green*, 411 U.S. 792 (1973).

8. See, e.g., *International Bhd. of Teamsters v. United States*, 431 U.S. 324 (1977). The term “pattern or practice” comes from § 707(a) of Title VII, which provides:

Whenever the Attorney General has reasonable cause to believe that any person or group of persons is engaged in a pattern or practice of resistance to the full enjoyment of any of the rights secured by this subchapter, and that the pattern or practice is of such a nature and is intended to deny the full exercise of the rights herein described, the Attorney General may bring a civil action . . .

42 U.S.C. § 2000e-6(a) (1988). Although the term “pattern or practice” in the statute refers to actions brought by the government, the term has been extended so that it often refers to a class action brought by private plaintiffs. The standards of proof for claims of systematic disparate treatment are the same whether the action is brought by the government or by private parties. *Cooper v. Federal Reserve Bank*, 467 U.S. 867, 876 n.9 (1984).

Opportunity Commission (EEOC)) must demonstrate that discrimination is the employer's standard operating procedure.

Statistical evidence has a role to play in each of the three forms of discrimination action described above—disparate impact, individual disparate treatment, and systematic disparate treatment. Statistical evidence is indispensable to a claim of disparate impact because the claim is that the challenged practice has an adverse effect on a group, not merely on an individual; that effect can be shown only by demonstrating statistically that the effects of the practice differ from group to group.⁹ Statistical evidence is useful, though not essential, in claims of individual disparate treatment, because the plaintiff's claim of adverse treatment based on race or sex may be buttressed by a showing that the employer routinely treats people of different groups differently. Conversely, the plaintiff's claim may be weakened by the employer's presentation of favorable statistics to diminish, but not eliminate, the probability that the plaintiff had been subjected to invidious discrimination.¹⁰ Finally, statistical evidence plays a central role in demonstrating that an employer has engaged in a pattern or practice of intentional discrimination, although, depending upon the particular court, such evidence may be neither necessary nor sufficient for a plaintiff to prevail.¹¹ It is primarily the use of statistical evidence of systematic intentional discrimination that is the focus of this Article, although much of the discussion is relevant to statistical proof in the other kinds of cases.

The centerpiece of a plaintiff's proof in the typical pattern-or-practice case is a demonstration that the observed representation of women or minorities in the employer's work force is lower than the representation that would be "expected" if employment decisions were made randomly with respect to race or sex; that demonstration is, in turn, coupled with an inference that underrepresentations are a consequence of intentional discrimination.¹² For example, if an employer has a

9. See WALTER B. CONNOLLY, JR. ET AL., *USE OF STATISTICS IN EQUAL EMPLOYMENT OPPORTUNITY LITIGATION* § 2.01[1] (1992) ("Disparate impact is, by definition, established by statistics since impact is described by quantitative patterns.").

10. See *Furnco Constr. Corp. v. Waters*, 438 U.S. 567, 579–80 (1978); *McDonnell Douglas*, 411 U.S. at 805.

11. See *Pitre v. Western Elec. Co.*, 843 F.2d 1262, 1267 (10th Cir. 1988) ("Although statistics may be useful to show differences in treatment and to establish a pattern and practice, they are clearly not required, especially when the sample size is too small to produce meaningful results.").

12. A finding of the existence of a pattern or practice of discrimination has the significant consequence of establishing a presumption that each individual class member was a victim of discrimination. *Franks v. Bowman Transp. Co.*, 424 U.S. 747, 772 (1976).

work force of 200 entry-level unskilled employees and blacks constitute 20 percent of the local labor force, the statistical "expectation" is that the nondiscriminating employer would have 40 black employees. Of course, because of random variations, one would not have any confidence that a nondiscriminating employer would have precisely 40 black employees, no more and no less. That is, no one would seriously suggest that the fact that the employer had 39 or 41 blacks demonstrates discrimination either for or against blacks,¹³ any more than one would conclude that failure to flip exactly 50 heads out of 100 coin tosses demonstrates a flaw in the coin. But what if the employer has 30 blacks instead of the expected 40? What if it has 28?

In order to distinguish those statistical disparities that are explainable on the basis of chance from those that are not, courts apply probability theory to test for the likelihood that a distribution as extreme as that observed would occur by chance if whites and blacks had a probability of being selected that was proportional to their representation in the applicant pool.¹⁴ In the example given above, most courts would use the normal approximation of the binomial distribution to assess the probability of so large a deviation.¹⁵ Applying this method to the employer with 30 blacks, it turns out that the standard deviation is 5.66,¹⁶ and 30 blacks corresponds to a z-score of 1.77,¹⁷

13. See Arthur B. Smith, Jr. & Thomas G. Abram, *Quantitative Analysis and Proof of Employment Discrimination*, 1981 U. ILL. L. REV. 33, 39 ("Labor economists . . . would not expect even a nondiscriminatory employer to at all times hire minorities and nonminorities at a rate exactly equal to their representation in the labor market.").

14. *E.g.*, Barnes v. GenCorp, 896 F.2d 1457 (6th Cir.), *cert. denied*, 111 S. Ct. 211 (1990); Palmer v. Shultz, 815 F.2d 84, 91 (D.C. Cir. 1987); Maddox v. Claytor, 764 F.2d 1539 (11th Cir. 1985).

15. See MICHAEL O. FINKELSTEIN & BRUCE LEVIN, *STATISTICS FOR LAWYERS* 120-22 (1990). The standard deviation is calculated using a formula

$$\sqrt{np(1-p)}$$

where n is the number of trials, and p is the probability of obtaining the result on a given trial.

16. Thus, in this case the standard deviation is obtained

$$\sqrt{(200 \times .20) \times (1 - .20)}$$

or

$$\sqrt{40 \times .80} = \sqrt{32} = 5.66$$

17. The z-score is the number of standard deviations above or below the mean. Louis J. Braun, *Statistics and the Law: Hypothesis Testing and Its Application to Title VII Cases*, 32 HASTINGS L.J. 59, 73 n.73 (1980). Here, the calculation is as follows:

$$z = \frac{\text{observed-expected}}{\text{s.d.}}$$

$$z = \frac{30-40}{5.66} = -1.77$$

which in turn corresponds to a probability of about 8 percent. That is, over the course of a large number of trials, if 200 employees were selected at random from a large pool of potential employees that was 20 percent black, a deviation this great would be obtained approximately 8 percent of the time. Because courts generally do not consider probabilities of more than 5 percent "statistically significant," most courts would hold that in this example the statistical evidence standing alone does not constitute a *prima facie* case of discrimination.¹⁸

A court might reach the opposite result if the employer had only 28 blacks. Twenty-eight blacks would correspond to 2.12 standard deviations below the expected value, with an associated probability of less than 3.5 percent. Using a 5-percent significance level, a court might find that this disparity is sufficiently unlikely to occur by chance that chance can be eliminated as an explanation for the disparity. Many courts recognize that to exclude chance as a factor is not the same as finding that impermissible discrimination has occurred;¹⁹ other courts are not so careful.²⁰ Whether or not they recognize this distinction, however, courts will often find such a showing sufficient to establish a *prima facie* case of discrimination, at which point the employer will

18. BARBARA SCHLEI & PAUL GROSSMAN, *EMPLOYMENT DISCRIMINATION LAW* 325 (2d ed. Supp. 1983-1985) (stating that most courts find that disparities of less than two standard deviations are not statistically significant); *see also infra* note 35.

19. *See, e.g., Barnes*, 896 F.2d at 1468-69:

When a plaintiff demonstrates a significant statistical disparity in the discharge rate, he or she has provided strong evidence that chance alone is not the cause of the discharge pattern. . . . [but t]he statistics do not and cannot determine whether the more likely cause is the defendant's bias or a legitimate selection criterion.

See also Palmer, 815 F.2d at 91 ("Nor can statistics determine, if chance is an unlikely explanation, whether the more probable cause was intentional discrimination or a legitimate nondiscriminatory factor in the selection process."); *Maddox*, 764 F.2d at 1552 ("It is important to stress that a disparity translating into a large number of standard deviations does not automatically point to discrimination as the cause."); *Gay v. Waiters' and Dairy Lunchmen's Union, Local 30*, 694 F.2d 531, 553 (9th Cir. 1982) ("[S]tatistics demonstrating that chance is *not* the more likely explanation are not by themselves sufficient to demonstrate that race *is* the more likely explanation for an employer's conduct.").

20. *See Capaci v. Katz & Besthoff, Inc.*, 711 F.2d 647, 651-52 (5th Cir. 1983) (employing standard-deviation analysis to compute the "probability of unbiased hiring"), *cert. denied*, 466 U.S. 927 (1984); *Payne v. Travenol Lab., Inc.*, 673 F.2d 798, 820 n.32 (5th Cir.) ("Absent explanation, standard deviations of greater than three generally signal discrimination . . ."), *cert. denied*, 459 U.S. 1038 (1982); *Rivera v. City of Wichita Falls*, 665 F.2d 531, 545 (5th Cir. 1982) ("Standard deviation analysis quantifies the likelihood of a benign explanation for an observed discrepancy . . ."); *Ivy v. Meridian Coca-Cola Bottling Co.*, 641 F. Supp. 157, 165 (S.D. Miss. 1986) ("A fluctuation of two or three standard deviations indicates that the result is caused by discriminatory intent rather than chance."); *see also John M. Dawson, Are Statisticians Being Fair to Employment Discrimination Plaintiffs?*, 21 *JURIMETRICS J.* 1, 8 (1980) (noting that a significance level of .05 "means tolerating a 5 percent chance . . . of a false inculpation of an employer in a Title VII case").

bear the burden of either attacking the *prima facie* case or demonstrating a nondiscriminatory explanation for the disparity.

The leap from statistical disparity to a *prima facie* case of discrimination is based on two powerful, necessary, but incorrect, assumptions—one logical and the other factual. The logical assumption, referred to here as the “Statistical Fallacy,” is that, in the example above, the probability of drawing 28 or fewer blacks at random—which we have already seen is approximately 3.5 percent—is the same as the probability that an observed work force having 28 or fewer blacks resulted from chance. The first probability is the probability of the observed disparity given random selection; the second is the probability of random selection given the observed disparity. As will be discussed below, there is no direct relationship between these two probabilities,²¹ yet courts uniformly interpret the first probability—which is the only one they have—as if it were evidence of the second, which is the one they need. Although in theory this logical flaw can inure to the benefit of defendants as well as plaintiffs—since it may lead courts to conclude that there is no discrimination when the disparity is not statistically significant²²—in practice this misunderstanding leads disproportionately to erroneous conclusions in favor of plaintiffs because of the disproportionate weight given to statistically significant results.

The factual assumption upon which the *prima facie* case of discrimination is founded, referred to here as the “Central Assumption,” is that in the absence of discrimination one would expect that the work force of each employer would—with only chance variations—mirror

21. See generally David H. Kaye, *Statistical Significance and the Burden of Persuasion*, LAW & CONTEMP. PROBS., Autumn 1983, at 13.

22. Compare *Ottaviani v. State Univ.*, 875 F.2d 365, 371 (2d Cir. 1989) (“Before a deviation from a predicted outcome can be considered probative, the deviation must be ‘statistically significant.’”), *cert. denied*, 493 U.S. 1021 (1990) with *Catlett v. Missouri Highway & Transp. Comm’n*, 828 F.2d 1260, 1265 (8th Cir. 1987) (“The inadequacy of the class’ statistics by themselves to establish discrimination does not foreclose recovery . . .”), *cert. denied*, 485 U.S. 1021 (1988) and *Maddox*, 764 F.2d at 1552 (“Just as a standard deviation of two or three does not necessarily exclude legitimate causes other than chance, so a Z-value below that range does not necessarily exclude discrimination as the cause.”).

Courts have shown themselves quite willing to find discrimination even in the absence of statistically significant statistical evidence if other, nonstatistical evidence, is available. Their recognition of the fact that an absence of statistically significant disparities does not by itself prove the absence of discrimination is in some contrast to their apparent obliviousness in disparate-impact cases to the fact that, when it comes to considering the defendant’s “business necessity” defense, the lack of a statistically significant relationship between a challenged practice and job performance does not necessarily mean that no relationship exists. See, e.g., *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 425–36 (1975) (discussing job relatedness).

the racial and sexual composition of the relevant labor force.²³ Put another way, courts assume that blacks, whites, males, and females are all equally likely to be qualified and available for—and interested in—each job. Although courts take some qualification differentials into account by comparing only people having the minimum qualifications for the job—if such minimum qualifications exist and if relevant data are available—and, in some instances take interest differentials into account by basing the statistical comparisons on applicant-flow data²⁴—thereby limiting the comparison to those potential employees who have indicated an interest in the job—the measures they use are so crude as to be virtually useless. For example, the Supreme Court in *Hazelwood School District v. United States*²⁵ found appropriate a statistical comparison between the proportion of black school teachers in a school district in the suburbs of St. Louis and the proportion of black school teachers in the entire St. Louis area.²⁶ The assumption underlying such a comparison is that within the class of certified teachers the average qualifications of blacks and whites are equal and that blacks and whites would be equally interested in obtaining a job in the school district at issue. The difficulty with this sort of assumption in general is that not only is it almost always unproved, it is often both demonstrably false as a matter of fact and inconsistent with other important assumptions of discrimination law.

23. This assumption is traceable in large part to the Supreme Court's opinion in *International Brotherhood of Teamsters v. United States*, 431 U.S. 324, 339–40 n.20 (1977), in which it stated:

Statistics showing racial or ethnic imbalance are probative in a case such as this one only because such imbalance is often a telltale sign of purposeful discrimination; absent explanation, it is ordinarily to be expected that nondiscriminatory hiring practices will in time result in a work force more or less representative of the racial and ethnic composition of the population in the community from which employees are hired.

See also Marcy M. Hallock, *The Numbers Game—The Use and Misuse of Statistics in Civil Rights Litigation*, 23 VILL. L. REV. 5, 12 (1977–78) (“If the employer has not discriminated on the basis of age, the proportion of covered individuals who applied for jobs will be similar to the proportion of covered individuals who were hired.”).

24. See *EEOC v. Rath Packing Co.*, 787 F.2d 318, 337 (8th Cir.) (“[A]pplicant flow data is often the best indicator of the extent of an employer's discrimination . . .”), *cert. denied*, 479 U.S. 910 (1986); *Bazemore v. Friday*, 751 F.2d 662, 687 (4th Cir. 1984) (indicating preference for applicant-flow data), *aff'd in part, vacated in part*, 478 U.S. 385 (1986); *United States v. County of Fairfax*, 629 F.2d 932, 940 (4th Cir. 1980) (indicating preference for applicant-flow data), *cert. denied*, 449 U.S. 1078 (1981); *Green v. United States Steel Corp.*, 570 F. Supp. 254, 275 (E.D. Pa. 1983) (noting that “actual applicant flow data is the preferred yardstick”), *aff'd in part, vacated in part sub nom. Green v. USX Corp.*, 843 F.2d 1511 (3d Cir. 1988), *vacated*, 490 U.S. 1103 (1989). See generally CONNOLLY ET AL., *supra* note 9, § 4.03.

25. 433 U.S. 299 (1977).

26. The Court remanded for consideration of whether, because of its strong affirmative action efforts, the City of St. Louis itself should be excluded from the comparison. *Id.* at 312–13.

The power of these twin assumptions has led courts to answer two questions inappropriately. The first question is whether an observed disparity is caused by chance or nonchance factors. The Statistical Fallacy—with its equating of the significance level and the probability of chance being the cause of the disparity—has caused courts to rule out chance as a cause in circumstances where such a conclusion is unwarranted. The second question is whether a disparity that is believed to have been caused by nonchance factors was caused by systematic illegal discrimination or instead some other systematic, but benign, cause. The Central Assumption that race- and sex-blind hiring would lead to proportional representation has led courts to conclude that differences in qualifications and interest are probably not the cause of a disparity and therefore to place upon employers the burden of proving that the groups at issue are not equally qualified or interested. As a result, it has fallen upon the defendant to demonstrate that the variables contained in a plaintiff's statistical analysis are wrong, rather than on the plaintiff to demonstrate that they are right. However, if "discrimination" under Title VII means treating similarly situated persons differently, proving that persons are similarly situated is rightfully part of the plaintiff's case.

The remainder of this Article will attempt to refute the Statistical Fallacy and the Central Assumption and show how they have improperly skewed proof requirements in Title VII pattern-or-practice cases, followed by suggestions for how such cases should be litigated in the future.

II. THE STATISTICAL FALLACY

The core of the Statistical Fallacy is the belief that the probability that a given event occurred by chance is the same as the prior probability of observing the event in a random draw. Suppose, for example, that there is only a one in 100 chance that a husband and wife of a given height would have a biological child who grows up to be over six-feet tall. The Statistical Fallacy would cause us to conclude that when we observe that phenomenon there is a 99-percent chance that the child is not the biological offspring of both parents. If we know that the woman is the biological mother (for example, we were present at the birth), we should be quite certain that artificial insemination or adultery is responsible for this rare event.

Appreciation of the defect in statistical reasoning requires no expertise in statistical methods; it requires only common sense, as can be demonstrated by using a simple coin-tossing example. We know that the probability of obtaining heads on a given toss is 0.5 if the coin is

fair; thus, the expected number of heads out of 100 tosses is 50. However, we also know that most series of tosses will not produce exactly 50 heads.²⁷ If we want to determine whether the coin is fair, we might toss the coin 100 times to see what we get. If we obtain 100 heads we can be quite confident that the coin is not fair; if we obtain 60 heads we would be substantially less confident in concluding lack of fairness; and if we obtained 55 heads, even less so. If we obtained exactly 50 heads, we would conclude that the observed distribution provides no support for a conclusion that the coin is not fair, although it would be improper to conclude that the distribution proves that the coin *is* fair.²⁸

A statistician would use the technique of hypothesis testing to make probabilistic statements about the fairness of a coin.²⁹ The “null hypothesis” would be that the coin is fair; that is, the real probability of obtaining heads with this particular coin is exactly 0.5.³⁰ The “alternative hypothesis” would be that the coin is not fair; that is, the real probability of heads is not 0.5, although there may be no notion of what the true probability actually is.³¹ The goal of hypothesis testing is to reject the null hypothesis, since it is the alternative hypothesis that is the real subject of the inquiry, i.e., that the coin is not fair (or, in discrimination cases, that the employer does not select employees without regard to race or sex).³²

27. In fact, if we routinely obtained an exact 50/50 split, we would suspect that the results had been doctored. See CONNOLLY ET AL., *supra* note 9, § 11.12 (Observed data are suspicious if they “look too much like the ideal model.”).

A famous example of data that are “too good” is Gregor Mendel’s data concerning inheritance of traits in garden peas. R.A. Fisher’s statistical analysis of these data caused him to conclude that a bias in favor of Mendel’s expectation “seems to pervade the whole of the data.” R.A. Fisher, *Has Mendel’s Work Been Rediscovered?*, 1 ANNALS SCI. 115, 131 (1936). In fact, for a series of experiments, the likelihood of data conforming so closely to Mendel’s theoretical predictions was only .00007, leading Fisher to conclude that the data may have been “falsified so as to agree closely with Mendel’s expectations.” *Id.* at 131, 132; see also Walter W. Piegorsch, *Fisher’s Contributions to Genetics and Heredity, with Special Emphasis on the Gregor Mendel Controversy*, 46 BIOMETRICS 915 (1990).

28. For example, even if the coin is unbalanced in a way that will lead to a 60/40 head-tail split over a large number of tosses, it is still possible in a particular series to obtain a 50/50 distribution by chance.

29. See Mikel Aickin, *Issues and Methods in Employment Discrimination Statistics*, 26 JURIMETRICS J. 347, 352–55 (1986).

30. See Braun, *supra* note 17, at 68 (“A null hypothesis postulates equality rather than inequality, presuming that whatever statistically demonstrated inequality does exist is due to chance factors.”).

31. The alternative hypothesis postulates a difference between groups. *Id.*

32. However, it is important to note that hypothesis testing does not allow one to “prove” the null hypothesis; even if the experimenter obtained an exact 50/50 split in one series of tosses, he could not state with any assurance that the true probability was exactly 0.5.

Prior to the experiment, the statistician would set a "significance level," which is the probability level at which he is willing to reject the null hypothesis.³³ A commonly used significance level is .05, which means that the null hypothesis will be rejected if the distribution observed in the experiment is one that would be expected to occur less than 5 percent of the time if the null hypothesis is true. Having established the ground rules, the statistician now tosses the coin and obtains 60 heads and 40 tails. What can he say about the coin? Applying the normal approximation of the binomial distribution he finds that 60 heads is exactly two standard deviations above the expected value of 50. Two standard deviations is equivalent to a probability just under 5 percent.³⁴ Therefore, the experimenter is prepared to reject the null hypothesis that the coin is fair using a 5-percent significance level.

Where the courts have gone wrong in employment discrimination cases is in reasoning backward. They assume that if an employer's work force reflects disparities of two standard deviations there is only a 5-percent chance that the disparity is due to chance³⁵ and, concomitantly, a 95-percent chance that it is due to nonchance factors,³⁶ fac-

33. See David H. Kaye, *Is Proof of Statistical Significance Relevant?*, 61 WASH. L. REV. 1333, 1342-45 (1986).

34. The probabilities used throughout are "two-tailed" probabilities. That is, the alternative hypothesis is that there is some difference between the two groups (e.g., blacks are not hired at the same rate as whites), although the difference could run in either direction. See Richard Goldstein, *Two Types of Statistical Errors in Employment Discrimination Cases*, 26 JURIMETRICS J. 32, 45-47 (1985). Under a one-tailed test, the alternative hypothesis is that the difference runs in a particular direction (e.g., blacks are hired at a lower rate than whites). Equivalent deviations are more statistically significant under a one-tailed test than they are under a two-tailed test. *Id.* at 47. Courts to date have tended to favor the use of two-tailed tests. But see *Little v. Master-Bilt Prods., Inc.*, 506 F. Supp. 319, 333 (N.D. Miss. 1980).

35. Not all courts have used the 5-percent significance level as the benchmark, but it is the most commonly used level. See *Dalley v. Michigan Blue Cross/Blue Shield, Inc.*, 612 F. Supp. 1444, 1451 n.18 (E.D. Mich. 1985) ("Most courts and commentators have accepted the .05 level" as the measure of statistical significance.). But see *Watson v. Fort Worth Bank & Trust*, 487 U.S. 977, 995 n.3 (1988) ("We have emphasized the useful role that statistical methods can have in Title VII cases, but we have not suggested that any particular number of 'standard deviations' can determine whether a plaintiff has made out a prima facie case in the complex area of employment discrimination."); *Ottaviani v. State Univ.*, 875 F.2d 365, 372 (2d Cir. 1989) ("While appellants' argument that a finding of two standard deviations should be equated with a prima facie case of discrimination under Title VII is not without initial appeal, we are constrained to reject such a formal 'litmus' test for assessing the legitimacy of Title VII claims."), *cert. denied*, 493 U.S. 1021 (1990).

36. There are a number of ways that chance can operate to produce an imbalanced work force. First, simply by chance, the mix of actual applicants before the employer at the time of hiring may be nonrandom. For example, the group of white applicants may, by chance, happen to have better qualifications than the group of black applicants, even if the larger populations from which those applicants come have equal qualifications. Recognizing this fact, courts and commentators have expressed a general preference for applicant-flow statistics over labor-force statistics, except in those cases where applicant-flow statistics are either unavailable or deemed

tors that they believe the employer should be called upon to demonstrate. This conclusion is equivalent to a conclusion in the coin-tossing example above that there is only a 5-percent chance that the coin tossed by the statistician is fair. This is demonstrably false, as the illustration below will demonstrate.

Suppose each of 100 persons in a room flips a coin 100 times, and five of them obtain a split of 60/40 or greater, which is the statistically expected number. The reasoning employed by the courts in discrimination cases would yield a conclusion that for any one of those five persons there is only a 5-percent probability that the result was a random deviation from the expected, and, conversely, that there is a 95-percent chance that the observed distribution was caused by non-random factors.³⁷ But suppose we add another fact unknown to the statistician: prior to the coin toss, each of the coins was carefully measured and calibrated, and it was secretly determined that each coin was fair. Now, what is the probability that those tossers who obtained 60/40 splits used an unfair coin? Because of the additional information, we now know that the probability is zero, barring errors in calibration. However, because the kind of analysis that courts engage in does not take into account the prior probability that the coin is fair, which it would have to do in order to assign a probability to the observed disparities, the court's conclusion would still be that because the disparity equalled two standard deviations, the deviation is statisti-

unreliable. A comparison of actual hires to applicants in cases where chance has operated to skew the applicant pools would show no statistical disparity if applicant-flow data are used.

Alternatively, even if qualifications of the two groups in the pool of actual applicants are equivalent, chance differences may exist in whether applications of one group tended disproportionately to be submitted at a time when no vacancy existed. Finally, even if individual white and black applicants are equally qualified and the employer hires in a race-blind way, it still may hire a disproportionate number from one race. If, for example, 5 pairs of equally qualified black and white applicants apply sequentially for five individual openings, if the employer hires in a race-blind fashion, there is a 1 in 32 chance that all five will be white (and an equal chance that all five will be black) and an almost 1 in 5 chance that either 4 or 5 will be white. In a given work force, chance may be operating in one or all of the above-described ways to produce a skewed work force.

37. One could prove even more: if the probability of a given individual obtaining a split greater than or equal to 60/40 is .05, the probability of all five getting such a split is $(.05)^5$

or less than one chance in three million.

The absurdity of the conclusion can be even more powerfully demonstrated by considering events with higher probabilities. The probability of obtaining heads on one toss is 0.5; therefore, when it is observed that heads resulted from the toss, the conclusion should be that there is a 50-percent chance that the coin is not fair.

cally significant and therefore highly probative of the existence of a nonrandom cause.³⁸

In the discrimination context, the probability that an employer's work-force disparities are a consequence of chance is completely dependent upon a statistic which the courts never have: the likelihood of discrimination prior to making the employment decision.³⁹ Although one might attempt some estimate of the percentage of employers who engage in systematic discrimination, the estimate could be no more than the crudest approximation. Moreover, even if that statistic existed, its use would be improper, since the consequence would be to hold an employer liable simply on the ground that a certain percentage of employers are believed to discriminate on a systematic basis.⁴⁰ Presumably no one would employ such logic to argue that

38. The same reasoning applied by the courts would lead one to convict all lottery winners of fraud. If the probability that any given ticket is a jackpot winner is one in one million, then we ought to say that there is only a one in one million chance that the person turning up at the lottery office with the winning ticket has a genuine ticket.

39. For this reason, there has been substantial debate in the literature over whether an explicitly Bayesian approach to statistical evidence should be adopted, primarily in the area of identification evidence. See, e.g., Michael O. Finkelstein & William B. Fairley, *A Bayesian Approach to Identification Evidence*, 83 HARV. L. REV. 489 (1970); David Kaye, *The Laws of Probability and the Law of the Land*, 47 U. CHI. L. REV. 34 (1979); Laurence H. Tribe, *A Further Critique of Mathematical Proof*, 84 HARV. L. REV. 1810 (1971); Tribe, *supra* note 4, at 1361-62.

The subject of Bayesian theory is a controversial one in legal/statistical circles and beyond the scope of this Article. See Lea Brilmayer & Lewis Kornhauser, *Review: Quantitative Methods and Legal Decisions*, 46 U. CHI. L. REV. 116, 135 n.68 (1978). However, the essence of the technique is to assign a "prior probability" based upon nonstatistical evidence and then to compute a "posterior probability" based upon the combined statistical and nonstatistical evidence. See Michael O. Finkelstein & William B. Fairley, *A Comment on "Trial by Mathematics,"* 84 HARV. L. REV. 1801, 1802 (1971). Thus, it is a way to calculate the probabilistic effect of additional pieces of information. See Brilmayer & Kornhauser, *supra*, at 135 n.68. Such an approach does not hold substantial promise in the typical pattern-or-practice case where statistical evidence predominates, since the posterior probability is based almost entirely on the statistical evidence. See DAVID C. BALDUS & JAMES W.L. COLE, *STATISTICAL PROOF OF DISCRIMINATION* § 9.21 (1980) ("[A]t the current state of the art, Bayesian statistics do not, in our opinion, provide judicially acceptable methods for meeting critical presuppositions of the method."); David H. Kaye, *The Numbers Game: Statistical Inference in Discrimination Cases*, 80 MICH. L. REV. 833, 854 n.67, 854-55 (1982) (book review) [hereinafter Kaye, *Numbers Game*] (describing the "difficulty of institutionalizing Bayesian inference in the law," even in relatively straightforward cases like *Hazelwood* and concluding that "it would not provide that much additional guidance to the fact-finder"); Richard Lempert, *Statistics in the Courtroom: Building on Rubinfeld*, 85 COLUM. L. REV. 1098, 1101 (1985) ("[T]he Bayesian perspective is unlikely to transform the way statistical data is evaluated for litigation purposes at any time in the foreseeable future.").

40. See Brilmayer & Kornhauser, *supra* note 39:

If a plaintiff tries to prove discrimination in a property owner's refusal to rent him a house, he should not be allowed to introduce data from the market as a whole that shows discriminatory patterns in the neighborhood. The individual owner simply cannot be held accountable for the activities of others, just as he should likewise be unable to refute through

a simple showing by the Internal Revenue Service (IRS) that 51 percent of taxpayers cheat on their income taxes would justify the imposition of civil penalties against a taxpayer.

The standard statistical analysis in discrimination cases is one that would support findings of liability against thousands of nondiscriminating employers.⁴¹ Assuming for illustrative purposes that there is no discrimination in the workplaces under consideration and that the employer hires randomly—and even assuming that all employees and applicants are fungible—one in twenty statistical comparisons will result in a conclusion that the observed disparity was caused by nonchance factors, and every one of those conclusions will be wrong. More than 5 percent of employers are at risk, however, because for any given employer there can be statistical comparisons for each job, for each department, for the various racial, ethnic, and sex groupings, and for different time periods.⁴² All employers of significant size would likely have substantial numbers of jobs for which a statistical disparity could be demonstrated.⁴³ Yet the true probability

general market statistics a showing of discrimination. Even if a utilitarian overview of the situation shows that it would result in greater overall accuracy to grant the would-be tenant a remedy, we would consider the evidence irrelevant to the particular case.

Id. at 150 n.129 (citation omitted).

41. Many of the cases that will be discussed involve multiple regression analysis, which, in essence, does the same sort of thing as identifying the relevant labor market. Regression analysis attempts to weigh the various factors that influence selection and determine whether there are effects related to a protected status. See Thomas J. Campbell, *Regression Analysis in Title VII Cases: Minimum Standards, Comparable Worth, and Other Issues Where Law and Statistics Meet*, 36 STAN. L. REV. 1299, 1305–08 (1984); Franklin M. Fisher, *Multiple Regression in Legal Proceedings*, 80 COLUM. L. REV. 702, 705–20 (1980); Note, *Beyond the Prima Facie Case in Employment Discrimination Law: Statistical Proof and Rebuttal*, 89 HARV. L. REV. 387, 397–98 (1975) (“Multiple regression . . . determines not merely whether a particular factor influences the observed characteristic, for example, wage, but assigns a number approximating *how much* the value of any given factor is weighted in a formula estimating the actual wage of an individual.”)

42. See *Pegues v. Mississippi State Employment Serv.*, 699 F.2d 760 (5th Cir.) (finding a pattern or practice for one six month period out of the six- to seven-year period encompassed by the lawsuit), *cert. denied*, 464 U.S. 991 (1983).

43. A plaintiff is thus in a position to “mine” the data for statistical disparities and allege discrimination in selected job groups. The defendant cannot necessarily avoid liability by demonstrating that there were no significant disparities in comparable job classifications. See, e.g., *Palmer v. Shultz*, 815 F.2d 84, 98 (D.C. Cir. 1987) (“[I]f plaintiffs in a Title VII case claim discrimination in [assignments to a particular ‘high visibility’ position], it is no defense that the government did not discriminate against women . . . in assignments to five other ‘high visibility’ positions.”).

As Mikel Aickin has pointed out, modern computer programs applied to a sufficiently rich data set will yield some statistically significant results even if the data set was derived from a random number table. Aickin, *supra* note 29, at 359. Therefore, he advocates questioning experts about statistical runs that they performed that did not yield significant results and that do not serve as the basis of their testimony. See *id.* at 358–59; see also CONNOLLY ET AL., *supra* note 9, § 11.09, at 11-24 (“A sufficiently comprehensive investigation nearly always results in the

that the observed disparities were obtained by something other than chance is, by hypothesis, zero.

The objection to the hypothesis-testing approach is not that it yields only approximate results or even that it occasionally yields incorrect results. Even our system of criminal justice with its "beyond a reasonable doubt standard" does not require infallibility; our civil standard of "a preponderance of the evidence" requires much less. Rather, the objection to hypothesis testing is that it yields meaningless results. A court's conclusion that a difference between the employer's actual work force and the statistically predicted work force demonstrates that the employer's work force probably (but not certainly) is the product of nonrandom factors is logically flawed. This error in analysis has dramatic consequences, for it causes a court faced with a statistically significant disparity to reason, "I'm faced with a disparity that is very unlikely to have occurred by chance; this rare result is suspicious, and the employer ought to explain it," when it should be thinking, "The plaintiff has described statistics that would be true for thousands of nondiscriminating employers; if the plaintiff wants me to suspect discrimination, he'd better give me a lot more than that."

This concern about the judicial misconception of the meaning of statistical evidence is quite real. The approach of most courts has been that a statistically significant disparity establishes a *prima facie* case unless the plaintiff's statistical comparison is so flawed that the assumption of fungibility of employees or applicants cannot be accepted.⁴⁴ When courts label a plaintiff's evidence "statistically significant," they generally mean that chance has been effectively ruled out as the cause and that they must look elsewhere for an explanation. At that point, the argument that the disparity was caused by chance is no longer open to the defendant, who then must attempt to show either that there is in fact no disparity or that the existing disparity did

discovery of some imbalance even among employees treated absolutely equitably."); Robert Follett & Finis Welch, *Testing for Discrimination in Employment Practices*, LAW & CONTEMP. PROBS., Autumn 1983 at 171, 183 ("[A]nalysis of an individual practice in isolation can lead to a mistaken inference of discrimination if the practice that is scrutinized is chosen because it presents the statistics that are least favorable to an employer.").

The difficulty for defendants in attempting to raise a "data mining" defense is that even though it is a logically sound argument, it may appear as an attempt to argue that the employer should not be liable for discrimination in the position at issue because it did not discriminate in other positions. This is precisely how the defense was misinterpreted in *Palmer*.

44. See, e.g., *Palmer*, 815 F.2d at 108 (holding that the inference of discrimination raised by significant disparities compels a judgment for the plaintiff in the absence of evidence in rebuttal); see *infra* notes 177-83 and accompanying text (discussing *Bazemore v. Friday*, 478 U.S. 385 (1986)).

not result from illegal discrimination.⁴⁵ Yet, as we have seen, the mere fact that a disparity is “statistically significant” does not show that it was probably not a consequence of chance.

The flaw in the statistical reasoning employed by courts and by the expert statisticians upon whom they rely is both fundamental and irrefutable. It is a mistake made by appellate courts of most, if not all, of the circuits,⁴⁶ numerous district

45. Although courts often talk about these as being separate methods of rebuttal, in most cases they are the same. If, for example, the employer argues in a hiring case that the plaintiff misdefined the relevant comparative labor force because the plaintiff did not take into account a particular qualification, the defendant is saying both that if the proper comparative populations were used, there would be no disparity, and that the disparity identified by the plaintiff is due to the difference in qualifications between the two groups and not due to invidious discrimination.

46. *Barnes v. GenCorp, Inc.* 896 F.2d 1457 (6th Cir.), *cert. denied*, 111 S. Ct. 211 (1990); *Rendon v. AT&T Technologies*, 883 F.2d 388, 398 (5th Cir. 1989) (upholding finding of discrimination based upon statistical expert testimony “that there was less than a five percent chance that the difference between the promotion times of minority installers and white installers could occur by chance”); *Ottaviani v. State Univ.*, 875 F.2d 365 (2d Cir. 1989), *cert. denied*, 493 U.S. 1021 (1990):

A finding of two standard deviations corresponds approximately to a one in twenty, or five percent, chance that a disparity is merely a random deviation from the norm When the results of a statistical analysis yield levels of statistical significance at or below the 0.05 level, chance explanations for a disparity become suspect, and most statisticians will begin to question the assumptions underlying their predictions.

Id. at 371; *Palmer*, 815 F.2d at 91 (“A statistical analysis of a disparity in selection rates can reveal the probability that the disparity is merely a random deviation from perfectly equal selection rates.”); *id.* at 92 (“[T]he .05 level . . . indicates that the odds are one in 20 that the result could have occurred by chance.”) (quoting *Segar v. Smith*, 738 F.2d 1249 (D.C. Cir. 1984)); *Maddox v. Claytor*, 764 F.2d 1539, 1552 n.13 (11th Cir. 1985) (“[F]rom a Z-value of -1.96, it would be concluded that there was a one-in-twenty chance that the raw disparity resulted solely through random distribution.”); *Coates v. Johnson & Johnson*, 756 F.2d 524, 537 n.13 (7th Cir. 1985) (stating that a p-value below .05 indicates “there is less than a 5% probability that the disparity was due to chance”); *Craik v. Minnesota State Univ. Bd.*, 731 F.2d 465 (8th Cir. 1984):

Statistical significance is a measure of the probability that an observed disparity is not due to chance. A finding that a disparity is statistically significant at the 0.05 or 0.01 level means that there is a 5 per cent. or 1 per cent. probability, respectively, that the disparity is due to chance.

Id. at 475 n.13 (citations omitted); *Lilly v. Harris-Teeter Supermarket*, 720 F.2d 326, 336 (4th Cir. 1983) (stating that disparities of more than two or three standard deviations “conclusively ruled out chance as the cause of the disparity in the termination rates”), *cert. denied*, 466 U.S. 951 (1984); *Carpenter v. Stephen F. Austin State Univ.*, 706 F.2d 608, 624 (5th Cir. 1983) (“Standard deviation analysis is a quantification . . . applied to statistical disparities to eliminate chance as a likely explanation for the revealed difference between an expected outcome and the observed outcome”); *Pegues*, 699 F.2d at 768 n.9 (“[T]he standard deviation quantifies the probability that chance is responsible for any difference between an expected outcome and the observed outcome”); *Payne v. Travenol Lab., Inc.*, 673 F.2d 798, 821 n.32 (5th Cir.) (“The ‘standard deviation’ is a way to calculate the likelihood that chance is responsible for the difference between a predicted result and an actual result.”), *cert. denied*, 459 U.S. 1038 (1982); *Rivera v. City of Wichita Falls*, 665 F.2d 531, 545 n.22 (5th Cir. 1982) (“A variation of two standard deviations would indicate that the probability of the observed outcome occurring purely by chance would be approximately five out of 100; that is, it could be said with a 95% certainty that

courts,⁴⁷ statistical expert witnesses,⁴⁸ and both legal and statistical

the outcome was not merely a fluke.") (citation omitted); EEOC v. American Nat'l Bank, 652 F.2d 1176 (4th Cir. 1981), *cert. denied*, 459 U.S. 923 (1982):

[S]tandard deviations can be expressed in terms of the mathematical probability that chance is the cause of the disparities. . . . Just short of two standard deviations—specifically at 1.96—the probability of chance is only 5 in 100; at just over two and one half, it is only 1 in 100; by three it is less than 1 in 100.

Id. at 1191–92 (citation omitted); Contreras v. City of Los Angeles, 656 F.2d 1267, 1273 n.3 (9th Cir. 1981) ("[A] .05 level of statistical significance means that the disparate results of a pre-employment screening device would be the product of chance only one time in twenty."), *cert. denied*, 455 U.S. 1021 (1982); Taylor v. Teletype Corp., 648 F.2d 1129, 1133 n.8 (8th Cir.) (stating that a disparity of 2.28 standard deviations "indicates statistically a likelihood of only one or two percent that pure chance was responsible for the demonstrated racial disparity"), *cert. denied*, 454 U.S. 969 (1981).

47. See, e.g., United States v. Lansdowne Swim Club, 713 F. Supp. 785, 809 (E.D. Pa. 1989) ("As a general rule, a standard deviation of greater than two or three excludes chance as an explanation for the underrepresentation of blacks."), *aff'd*, 894 F.2d 83 (3d Cir. 1990); EEOC v. Andrew Corp., 51 Empl. Prac. Dec. (CCH) ¶39,364 at 59,540 (N.D. Ill. Sept. 12, 1989) (stating that three standard deviations "is merely a shorthand for indicating the probability that Andrew's hiring pattern with regard to black clericals happened by chance"); Harrell v. University of Montevallo, 673 F. Supp. 430, 433 (N.D. Ala. 1987) ("The probability that the above described racial disparities in the statistics resulted from chance is extremely small—only one out of 100 times."), *aff'd*, 861 F.2d 725 (11th Cir. 1988); EEOC v. Sears, Roebuck & Co., 628 F. Supp. 1264, 1286 (N.D. Ill. 1986), *aff'd*, 839 F.2d 302 (7th Cir. 1988) ("A standard deviation of 2 indicates that there is a .045 probability (that is, almost a 5% probability), that the results observed occurred due to chance."); Rossini v. Ogilvy & Mather, Inc., 597 F. Supp. 1120, 1161 (S.D.N.Y. 1984) ("[A] t-statistic of approximately two indicates that the chances are less than one in 20 that the gender coefficient is the product of chance factors and the true coefficient of gender is zero."); Woodard v. Lehman, 530 F. Supp. 139, 145 (D.S.C. 1982) ("The plaintiff [can] demonstrate that a particular racial composition could not have occurred by chance . . . only where the difference between the expected black representation and the actual result exceeds two or three standard deviations."), *aff'd*, 717 F.2d 909 (4th Cir. 1983); Chisholm v. United States Postal Serv., 516 F. Supp. 810, 826 (W.D.N.C. 1980) ("[The] result could not have occurred by chance more than one time in a thousand. It is probable almost to the point of certainty that something other than the ability the test is supposed to measure is responsible for the racially differential results."), *aff'd in part, vacated in part*, 665 F.2d 482 (4th Cir. 1981); Gay v. Waiters' & Dairy Lunchmen's Union, Local No. 30, 489 F. Supp. 282, 308–09 n.33 (N.D. Cal. 1980) ("By reference to tables based on the normal distribution curve, the Z statistic can be used to determine the probability that the observed value is the product of random selection or occurrence."), *aff'd*, 694 F.2d 531 (9th Cir. 1982).

Judge Higginbotham's epic opinion in *Vuyanich v. Republic National Bank*, 505 F. Supp. 224 (N.D. Tex. 1980), *vacated*, 723 F.2d 1195 (5th Cir.), *cert. denied*, 469 U.S. 1073 (1984), noted in passing that "[a] test of statistical significance does not determine the probability that any particular result in fact occurred by chance," but there is no indication in his opinion that this observation played any role in his analysis of the statistical evidence. *Id.* at 348 (citing *BALDUS & COLE*, *supra* note 39). Indeed, he refers to disparities significant at the 5-percent level as "apparently discriminatory results," without any discussion of why they would appear so if the 5-percent level is not the probability of chance occurrence. *Id.* at 347.

48. *Mister v. Illinois Cent. Gulf R.R.*, 832 F.2d 1427, 1429 (7th Cir. 1987) ("The plaintiffs' expert witness testified that there was less than one chance in a million that this disparity was consistent with race-neutral hiring."), *cert. denied*, 485 U.S. 1035 (1988); *Blum v. Witco Chem. Corp.*, 829 F.2d 367, 371 (3d Cir. 1987) ("[P]laintiffs produced a statistical expert who testified that the probability that the disparate retention rate was due to some random factor unrelated to

commentators.⁴⁹ Although a few commentators have powerfully

age was .0084.”); *Moore v. McGraw Edison Co.*, 804 F.2d 1026, 1031 (8th Cir. 1986) (plaintiffs’ expert “testified that the chances were two in one thousand that age was not a factor in the terminations”); *Lewis v. Bloomsburg Mills, Inc.*, 773 F.2d 561, 569 (4th Cir. 1985) (“Unchallenged expert opinion testimony established that [a disparity in the range of 5-8 standard deviations] was statistically significant in the sense that the probability that it might be explained by chance alone was no better than one in a thousand.”); *Association Against Discrimination in Employment, Inc. v. City of Bridgeport*, 647 F.2d 256, 264 n.7 (2d Cir. 1981) (“Plaintiffs’ expert testified that the likelihood that such a difference would be attributable to chance would be less than five chances in one thousand.”), *cert. denied*, 455 U.S. 988 (1982); *Chance v. Board of Examiners*, 330 F. Supp. 203, 212 (S.D.N.Y. 1971), *aff’d*, 458 F.2d 1167 (2d Cir. 1972) (expert testified that “the probability of the difference being a chance result not related to the factor of race is determined as *less than one in one billion*”).

49. SCHLEI & GROSSMAN, *supra* note 18, at 1370–71 (2d ed. 1983):

Since chance can never be totally ruled out as a cause of any statistical disparity, the question becomes what risk is the court willing to assume that the disparity was caused by chance rather than an alternate explanation such as discrimination. . . .

The question then becomes what test of statistical significance will be used.

See id. at 1372 (“The .05 level of statistical significance means that the probability of the statistical disparity occurring by chance is 5 percent, which is to say one chance out of 20.”); CHARLES A. SULLIVAN ET AL., *FEDERAL STATUTORY LAW OF EMPLOYMENT DISCRIMINATION* § 1.8, at 74 (1980) (A disparity greater than 1.96 standard deviations “could be consistent with the [null] hypothesis less than 5% of the time. Rejecting the null hypothesis means that it is much more likely than not (though not certain) that the null hypothesis is not true.”); Anthony E. Boardman & Aidan R. Vining, *The Role of Probative Statistics in Employment Discrimination Cases*, *LAW & CONTEMP. PROBS.*, Autumn 1983, at 189, 205–06 (“To say that a relationship is statistically significant at the 95% level of confidence means that there is a 5% or lower probability of concluding that there is a relationship when, in fact, there is none.”); Braun, *supra* note 17, at 87 (“When led to a rejection of the null hypothesis at a level of significance of 0.05, a court can be at least 95% confident that a disparity of treatment of the relevant groups exists.”); Campbell, *supra* note 41, at 1306–09 (assuming that where the statistical model accurately reflects the employment process, the significance level reflects the likelihood that a finding of liability would be erroneous); Dawson, *supra* note 20, at 1–2 (“As applied in employment discrimination disputes, this 5 percent test standard represents a sanctioned risk of false inculcation of a defendant-employer, and is the burden of persuasion that must be met by a plaintiff who asserts that the rates differ because of unlawful discrimination.”); *id.* at 8 (“If we reject the null hypothesis in a test at alpha level A, our ‘confidence’ that this decision is right is 100 percent—A (e.g., 95 percent when alpha is 5 percent).”); Goldstein, *supra* note 34, at 40 (“[W]hen there is no difference in the population and we set [the significance level at] .05, . . . there is a five percent chance that we will incorrectly say that there is a difference between the two groups.”); Hallock, *supra* note 23, at 13 (“[T]he use of the .05 level of significance in employment discrimination cases seems appropriate [because] most people consider the one-out-of-twenty chance that a conclusion is wrong to be an acceptable degree of risk”); Howard C. Hay, *The Use of Statistics to Disprove Employment Discrimination*, 29 *LAB. L.J.* 430, 433 (1978) (“[E]fforts to define ‘substantial disparity’ are in fact simply efforts to determine whether the statistical disparity in question is attributable to ‘chance’ or ‘not to chance.’”); Jan W. Henkel & Patrick G. McKeown, *Unlawful Discrimination and Statistical Proof: An Analysis*, 22 *JURIMETRICS J.* 34 (1981):

Usually, if the probability of an outcome occurring is less than some predetermined level, say 0.05, and that outcome does occur, then the occurrence is said to be statistically significant. This does not mean that the outcome won’t occur, but it does imply that its occurrence cannot be explained by random variation.

Id. at 37; Barbara A. Norris, *Multiple Regression Analysis in Title VII Cases: A Structural Approach to Attacks of “Missing Factors” and “Pre-Act Discrimination,”* *LAW & CONTEMP.*

pointed out the defect in logic,⁵⁰ courts have largely ignored their efforts, and in some cases it appears that some of these commentators themselves do not fully understand the significance of their observations.⁵¹ This lapse in logic is difficult to explain on the basis of bias in

PROBS., Autumn 1986, at 63, 65–66 (explaining that results significant at the five percent level means that “the likelihood that they occurred by chance is no more than one in twenty”); Elaine W. Shoben, *In Defense of Disparate Impact Analysis Under Title VII: A Reply to Dr. Cohn*, 55 IND. L.J. 515 (1980):

The level of ‘statistical significance’ refers essentially to how certain an investigator wants to be that a difference found in a sample might not have occurred by chance alone even if there is no difference in the unknown population. Social scientists often adopt a five percent level of significance, meaning that there is no more than one chance in twenty that the difference appearing in the sample could have happened by chance alone even if there is no difference in the population.

Id. at 523 n.38; Smith & Abram, *supra* note 13, at 41 n.31 (“A hire rate disparity is deemed to be ‘significant’ if the probability that the observed disparity was due to chance is lower than a certain percentage level.”).

50. *E.g.*, Kaye, *supra* note 21:

The court’s assumption, however, that when the “probability of [statistical] error is less than 5%,” the “scientific fact is at least 95% certain” exemplifies a common misunderstanding of the role of statistical tests in scientific inference. . . .

. . . .

The difficulty is that this interpretation of the result of the hypothesis test is wrong. The test was structured so as to retain the null hypothesis unless the chance of getting the evidence under this hypothesis fell below 5%. The test focused exclusively on the probability of the evidence given the null hypothesis. Nothing was said about the probability of the hypothesis in the light of the experimental evidence. It may be tempting to call the probability of 0.055 the chance of a coincidence, and to say that the probability of something other than a coincidence — of foul play—must be what is left over, namely 0.945.

But this only shows that one can “prove” anything with words.

Id. at 21–22; see also Paul Meier et al., *What Happened in Hazelwood: Statistics, Employment Discrimination and the 80% Rule*, 1984 AM. B. FOUND. RES. J. 139, 143 (pointing out that “[i]t is a serious but very frequent error” to confuse the probability of random selection with the p-value). Follett and Welch have made the same point:

It is important to recognize that the 0.05 rule does not say that when a difference as large as two standard deviations occurs the probability that the two groups are treated equally is 5% or less, nor does it say that the probability of unequal treatment is 95% or more. It only says that if the two groups are treated equally, one would expect a difference as large as the one observed 5% of the time.

Follett & Welch, *supra* note 43, at 174.

51. A leading treatise on statistical proof of discrimination, for example, points out:

[E]ven if such a result would happen only 1 out of 100 times in a purely random system, a statistical test cannot measure the likelihood that a particular outcome was or was not a chance result. The test of significance states the likelihood of seeing evidence of the type observed over the long run, if the system were effectively blind to [any form of bias such as] color or sex. It simply does not follow that the chances are 99 out of 100 that the result in the instant case was not a chance result, let alone intentionally caused.

BALDUS & COLE, *supra* note 39, § 9.42. Although recognizing this essential point, the authors also opine:

[T]he statistical test can say nothing about causation. The [statistical] test . . . provide[s] relevant evidence on which an inference about causation can properly be based. For, if the observed result would occur by chance only 1 time out of 100 in a random system, the

favor of plaintiffs, since it infects both defense experts and courts that ultimately rule in favor of defendants.⁵²

In light of the fundamentally faulty statistical inferences this mode of analysis induces, it is remarkable that several commentators have argued not that using hypothesis testing with a 5-percent significance level is too easy on plaintiffs, but rather that it is too hard. Thus, a number of commentators have suggested that a 5-percent significance level requires plaintiffs to prove their case to a 95-percent certainty,

occurrence of the result observed in the case before the court would have to be considered a rare event (unless the system were not random), thereby suggesting chance was not the cause. And if bias is the only other plausible explanation for the result, one's belief that the defendant intentionally discriminated will be strengthened.

Id. § 9.42. However, as the prior discussion reveals, to say that the observed result would occur by chance only one time out of one-hundred merely tells one that the statistical pattern is not inconsistent with systematic discrimination; it in no sense *suggests* it. Baldus and Cole further advance the Statistical Fallacy by asserting that "[t]he risk of an erroneous rejection of the hypothesis that the disparity was caused by chance is considered a 'Type I' error, i.e., rejecting the null hypothesis when it should have been accepted," thus equating the probability of a random cause with the significance level. *Id.* § 9.0 n.5.

Even David Kaye, who has pointed out the Statistical Fallacy most forcefully, slips into language that in effect incorporates it. For example, he describes the significance level as the chance of a "false alarm," that is, the probability of falsely finding that a disparity was caused by nonchance factors when in fact it was caused by chance. Kaye, *Numbers Game*, *supra* note 39, at 844 n.41. He states:

In always rejecting the null hypothesis of no discrimination at a significance level of $\alpha = .05$, the Court will erroneously reject this hypothesis five percent of the time, since five percent of the cases . . . will occur due to chance alone, and the null hypothesis asserts the existence of such chance results.

Id. Kaye's statement is true only if employers are sued randomly with respect to the statistical profile of their work forces—i.e., if employers with a statistical overrepresentation of the relevant group and those with a statistically "expected" representation are as likely to be sued for systematic discrimination as employers with a statistical underrepresentation. If most employers coming before a court were selected *because* of their statistical disparities, the probability of error is much higher than—and not directly related to—the probability suggested by the significance level.

52. In the criminal context, courts and commentators have been quick to recognize the Statistical Fallacy. In the notorious case of *People v. Collins*, 438 P.2d 33 (Cal. 1968), the California Supreme Court reversed criminal convictions that were based upon the logic of the Statistical Fallacy. In that case, witnesses testified that they had seen a blond woman with a ponytail and a black man with a beard and mustache drive away in a yellow car immediately following a bank robbery. The defendants met that description. The prosecution introduced an expert witness to testify that the probability that a randomly chosen couple would match this description was one in twelve million, and the defendants were convicted. The court found several flaws in the statistical presentation, but the one of interest here is that it recognized that the probability that a random couple would match the description is entirely different from the probability that a couple matching the description is innocent. *Id.* at 40–41. Numerous scholars have commented, virtually all favorably, on the court's reversal and its statistical analysis. See Neil B. Cohen, *Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge*, 60 N.Y.U. L. REV. 385, 390 n.33 (1985) (citing 25 articles inspired by "the abuse of probability theory in *Collins*").

rather than by a mere preponderance of the evidence, which is assumed to be 51 percent.⁵³ However, as David Kaye has pointed out, there is no simple relationship between the significance level and the burden of persuasion.⁵⁴ If there were, a 5-percent significance level would probably come close to satisfying the criminal standard of proof, or at least the "clear and convincing" standard.

Recognition of the lack of clear relationship between the significance level and the probability that observed disparities in a given employer's work force were obtained by chance means that a plaintiff who demonstrates a disparity of two or three standard deviations has shown very little, since such showings could be made against most employers for some job categories unless the employer is systematically employing quotas. Since the *prima facie* case should exclude the most likely causes of a statistical disparity, such a weak statistical showing should not be taken as establishing a *prima facie* case.

One's immediate reaction might be that results that are significant at the 5-percent level, even if not determinative, are sufficiently rare that they are at least probative of discrimination.⁵⁵ However, contrary

53. See, e.g., Braun, *supra* note 17:

[T]he courts appear more reluctant to blame an employer who is actually not guilty of discrimination (rejecting a true null hypothesis and making a Type I error with respect to the innocence of the employer defendant) than to exculpate a defendant who did discriminate unjustly.

....

... Thus, the plaintiff must establish that there is at least a 95% chance that the defendant's actions were discriminatory before the court will hold that a *prima facie* case has been established by a preponderance of the evidence.

Id. at 69-70; Dawson, *supra* note 20, at 2 (The five-percent significance level "may be unduly onerous, since a very low risk of false *inculpation* (as the 5 percent standard is) may be associated with a high risk of false *exculpation*."); Henkel & McKeown, *supra* note 49, at 44-46; Marcel C. Garaud, Comment, *Legal Standards and Statistical Proof in Title VII Litigation: In Search of a Coherent Disparate Impact Model*, 139 U. PA. L. REV. 455 (1990):

The 95% confidence coefficient cut-off that courts have adopted is a stringent criterion developed to test scientific conclusions reached on the basis of imperfect or incomplete data, ... [which requires plaintiffs] to assemble an analysis in which the risk of falsely inculcating a defendant, that is, Type I risk, is no more than 5%.

Id. at 468; see also *Ethyl Corp. v. EPA*, 541 F.2d 1 (D.C. Cir.), *cert. denied*, 426 U.S. 941 (1976):

Typically, a scientist will not so certify evidence unless the probability of error, by standard statistical measurement, is less than 5%. That is, scientific fact is at least 95% certain.

Such certainty has never characterized the judicial or administrative process. It may be that the "beyond a reasonable doubt" standard of criminal law demands 95% certainty. ... But the standard of ordinary civil litigation, a preponderance of the evidence, demands only 51% certainty.

Id. at 28 n.58.

54. See Kaye, *supra* note 33, at 1339 n.31. See generally Kaye, *supra* note 21, at 23.

55. See BALDUS & COLE, *supra* note 39, § 9.42.

to the view of some courts, rarity does not imply an unlawful cause.⁵⁶ Moreover, the fallacious nature of the initial assumption becomes more obvious when it is recognized that lawsuits alleging class-wide discrimination are not brought at random; instead, they are selectively targeted against employers whose work force is "out of balance." Not surprisingly, therefore, there are very few reported pattern-or-practice cases, whether they find for or against the defendant, where there are not statistical disparities that either are "statistically significant" or at least approach statistical significance. If defendants in pattern-or-practice cases are selected *because* of their statistical disparities, the assumption that randomly caused disparities will be rare is without foundation.⁵⁷ To pursue the coin-tossing example employed above, if there were a Department of Fair Coins charged with monitoring coin tosses, it would certainly concentrate its energies on the coin tossers who obtained statistically significant results, but, depending upon the prevalence of unfair coins, it may in most cases be wrongly accusing the tosser of using an unfair coin.⁵⁸

Richard Lempert, although acknowledging that employment discrimination lawsuits are not brought at random, asserts that non-randomness tends to *increase* the proportion of valid claims.⁵⁹ In responding to Daniel L. Rubinfeld,⁶⁰ who had made the point that one would expect plaintiffs disproportionately to sue companies having significant disparities rather than companies with more balanced work forces, Lempert argues that "[e]ven if companies hire independently, investigations into discrimination are not started, and are certainly not

56. A number of courts in tort cases, for example, have made the error of assuming that if an outcome is rare, it is likely due to negligence. *See, e.g.,* McGonigal v. Gearhart Indus., Inc., 788 F.2d 321, 327 (5th Cir. 1986) ("If a man with thirty years experience in the munitions field had not previously heard of a grenade with a missing delay column, then it also seems reasonable to conclude that such an error will occur only when the inspector is negligent.").

57. *See supra* note 50.

58. Analogies to statistical proof in discrimination cases arise in statistical identification cases. Suppose a criminal defendant is charged with rape and a blood sample is taken from him that shows a match with a forensic specimen recovered from the victim. The prosecution puts on an expert who testifies that the likelihood of a random match is one in 1,000. Whether this is strong evidence for the prosecution depends on how the defendant was selected as a suspect. If the defendant had met the victim's description, been arrested near the scene of the crime, and then been picked out of a line-up by the victim, the statistical evidence would be strong confirmation indeed. On the other hand, if the police had a data bank including blood data on the entire population and selected the defendant simply because of the forensic match, the statistical evidence would be quite weak, since there may be hundreds or thousands of people in that particular city whose blood would match that taken from the victim.

59. Lempert, *supra* note 39, at 1114.

60. Daniel L. Rubinfeld, *Econometrics in the Courtroom*, 85 COLUM. L. REV. 1048, 1064 (1985).

pursued to the point of a trial, at random.”⁶¹ That is true enough. However, he fails to appreciate that the nonrandomness skews the cases *against* defendants. On the contrary, he argues that “I expect that before a case reaches the point of a substantial investment in a careful statistical analysis (often at the expense of the plaintiff’s attorney), it appears to someone that there is a high prior probability that the defendant company has been discriminating.”⁶²

Lempert’s argument is, on close examination, tautological. To the extent that the appearance of a “high prior probability” of discrimination is based upon the statistical disparities, those disparities are, in effect, being double counted. That is, a plaintiff may be denied a job and, because of the statistical imbalance in the employer’s work force, assume that it was because of discrimination. Alternatively, the imbalance itself, rather than any individual adverse employment decision, may have prompted the lawsuit.⁶³ What causes plaintiffs’ lawyers to make a “substantial investment in a careful statistical analysis” is not the lawyer’s estimated high prior probability that the defendant is discriminating; instead, it is the estimated high probability of prevailing in a lawsuit, with its associated high probability of recovering statutory attorneys fees. Sometimes the two probabilities are related, but to the extent that the estimated probability of prevailing is inflated by a misunderstanding of the meaning of the statistics, a plaintiff’s lawyer’s belief that a case is a good investment cannot be taken as an indication that the employer is discriminating.

Assuming for now that observed disparities are due either to chance or to discrimination, rather than to differences in qualifications or interest, it is very difficult to determine in a given case whether an observed disparity is due to chance or to discrimination. As demonstrated above, courts have erred in assuming that the significance level reveals the probability that chance was responsible, an error that has caused them to assume that either a systematic difference in qualifications or systematic discrimination was responsible for a disparity and to require the employer to demonstrate a legitimate systematic cause.⁶⁴ However, if pure chance is the cause—where solely by chance

61. Lempert, *supra* note 39, at 1114.

62. *Id.*

63. See *EEOC v. Sears, Roebuck & Co.*, 839 F.2d 302 (7th Cir. 1988).

64. See, e.g., *Waisome v. Port Auth.*, 948 F.2d 1370, 1376 (2d Cir. 1991) (“Social scientists consider a finding of two standard deviations significant, meaning there is about one chance in 20 that the explanation for a deviation could be random and the deviation must be accounted for by some factor other than chance.”); *Berger v. Iron Workers Reinforced Rodmen Local 201*, 843 F.2d 1395, 1419 (D.C. Cir. 1988) (The possibility that the challenged disparity was due to chance was “eliminated as part of plaintiffs’ *prima facie* statistical case.”), *cert. denied*, 490 U.S. 1105

whites obtained jobs at a disproportionately high rate over equally qualified blacks⁶⁵—a search for a systematic explanation is likely to be fruitless, since by definition there is no systematic difference between the two groups and no systematic cause of the disparity.⁶⁶ If the courts take account of this fact, the analysis of the respective burdens of the parties must change.

Treating statistically significant differences as *prima facie* proof of discrimination—and thereby calling upon employers to prove a systematic, but nondiscriminatory, cause under pain of liability—places an insurmountable burden on many employers in those cases where the disparities are actually due to chance. Chance disparities that are significant at the 5-percent level are ubiquitous, but under current law highly suspect. Nondiscriminatory disparities are therefore very costly for employers, exacerbating the pressures on them to adopt either overt or surreptitious quotas.⁶⁷

Once it is understood that chance disparities are common and that statistical analysis does not reveal the probability that a disparity was a consequence of chance, the justification for treating disparities as *prima facie* proof of discrimination disappears. There are two primary reasons for treating statistical disparities as *prima facie* proof of dis-

(1989); *Warren v. Halstead Indus., Inc.*, 802 F.2d 746, 759 n.17 (4th Cir. 1986) (“[A] standard deviation greater than two or three necessarily excludes ‘chance’ as the cause of under representation.”); *Craik v. Minnesota State Univ. Bd.*, 731 F.2d 465, 475–76 n.13 (8th Cir. 1984) (“Normally, courts and social scientists, when confronted with such low probabilities, conclude that something other than chance is causing the disparity. The courts must decide whether that ‘something’ is illegal discrimination or some other, nondiscriminatory influence.”); *Lilly v. Harris-Teeter Supermarket*, 720 F.2d 326, 337 n.22 (4th Cir. 1983) (“[T]he standard deviation analysis demonstrates that the significant disparity between the termination rates was not due to chance, but instead to some other cause. That cause is presumed to be unlawful discrimination until the defendant shows otherwise . . .”), *cert. denied*, 466 U.S. 951 (1984); *EEOC v. American Nat’l Bank*, 652 F.2d 1176, 1192 (4th Cir. 1981) (“[W]ith standard deviations of more than three, the analysis may perhaps safely be used absolutely to exclude chance as a hypothesis, hence absolutely to confirm the legitimacy of an inference of discrimination based upon judicial appraisals that disparities are, to the legally trained eye, ‘gross.’”), *cert. denied*, 459 U.S. 923 (1982).

65. The converse is possible as well (i.e., chance could favor blacks), but a lawsuit is less likely to result from that.

66. If the disparity is caused by differences in timing, no evidence is likely to exist, because employers often do not keep records of which applications they have before them at a particular time.

67. As Judge Higginbotham wrote in *Vuyanich v. Republic National Bank*, 505 F. Supp. 224 (N.D. Tex. 1980), *vacated*, 723 F.2d 1195 (5th Cir.), *cert. denied*, 469 U.S. 1073 (1984):

We know the mathematical techniques also may be freighted with collateral consequences, some antithetical to the aims of the legislation being enforced. For example, we blink at reality when we say that ratios and quotas are not being reinforced when liability is heavily hinged upon “disparities” and deviations from numerical exemplars.

Id. at 394.

crimination under a legal regime that expressly eschews a proportional-representation requirement: either the evidence is highly suggestive of discrimination or the defendant's access to explanatory evidence is sufficient to require the defendant to come forward with it.⁶⁸ As already demonstrated, the first justification does not exist, since the pervasiveness of random disparities makes it impossible to say that a disparity, standing alone, is most likely a consequence of discrimination. The second justification is also inadequate. Where disparities are in fact a result of chance, the employer is unlikely to have evidence to support that fact, and in those situations where there are systematic differences between the relevant populations as a whole, that evidence is as available to the plaintiff as it is to the defendant.⁶⁹ If a burden of production is placed on the employer notwithstanding the fact that the evidence that the employer is required to produce is unlikely to exist, the procedural device of shifting the burden of production is converted into a substantive rule making the employer liable for chance disparities that exceed a given magnitude—i.e., chance disparities that are “statistically significant.” In short, the critical point is that statistical evidence in these circumstances is simply inconclusive and therefore cannot provide an adequate basis for determining liability. Once this point is recognized, appropriate rules on burdens of proof will readily follow.

A rule that has the consequence of rendering the employer liable for chance disparities seems inconsistent with a legal framework under which “[s]tatistics showing racial or ethnic imbalance are probative . . . only because [they are] often a telltale sign of purposeful discrimination”⁷⁰ Nonetheless, a system under which employers would expressly be liable for random disparities has its defenders. For example, Richard Lempert provides three reasons for holding employers liable for such disparities.⁷¹ First, he argues that simply showing that randomness *may* be the cause is not the same as showing that it probably *was* the cause.⁷² Although this may be true, it assumes that the burden of proof on the issue ought to rest on the defendant. However, unless we can say—which we cannot—that disparities are usually caused by discrimination rather than by chance, placement of the bur-

68. *See, e.g., International Bhd. of Teamsters v. United States*, 431 U.S. 324, 359 n.45 (1977).

69. Employers as a group have no special knowledge of general sociological tendencies, such as the aversion or attraction of various groups to certain kinds of jobs or of the relative frequencies of various qualifications within groups.

70. *Teamsters*, 431 U.S. at 339 n.20.

71. Lempert, *supra* note 39, at 1115–16.

72. *Id.* at 1115.

den on the defendant is unwarranted, and there is no reason to depart from ordinary rules of evidence that require the plaintiff to establish by a preponderance of the evidence that the defendant has violated the law.

Lempert's next argument is that whether disparities are caused by intent or by chance, we may wish to outlaw them to advance the interests of a group that has been subject to historic discrimination and therefore require that the company hire "disfavored" groups in proportion to their representation in the applicant pool.⁷³ This argument calls for a change in substantive law—i.e., a proportional-hiring rule—in the guise of tinkering with the rules of evidence.

Finally, and perhaps most troubling, Lempert suggests a fundamentally lawless approach, arguing that we may want to make an example out of even non-discriminating employers both to demonstrate how much we loathe discrimination and to punish discriminating employers who would not be held liable under a regime that required evidence.⁷⁴ Lempert suggests that this "value conflict" between imposing or not imposing liability for random disparities might be resolved either way.⁷⁵ However, the values that conflict are not commensurate. The value of requiring evidence before holding a defendant liable and the value of not holding employers liable for random statistical disparities are values that emanate from our basic legal traditions and from Title VII, respectively.⁷⁶ The values furthered by imposing liability simply to make the points suggested by Lempert are Lempert's own values, values that a court ought not consider in resolving the issue.

Lempert's suggestions illustrate an unfortunate tendency to disregard ordinary notions of fairness when issues of equality are impli-

73. *Id.*

74. *Id.* at 1115–16. Lempert writes:

Third, for purposes of deterrence we may not want to allow a company with a disproportionately small number of women employees to appear untouched by the legal process. Other companies with a propensity to discriminate may thereby be emboldened to do so. Moreover, to allow the defense of randomness may mean that companies that are in fact discriminating escape by using it, thus doing injustice in particular cases and lowering specific as well as general deterrence.

Id.

75. *Id.* at 1116.

76. In light of the incompatibility of Lempert's suggestions with applicable law, it is somewhat surprising that he prefaced his discussion of random effects as follows:

If there is any general theme to my comments it is *Lex Regis*: when econometrics or any other statistical specialty enters the courtroom, the law is king. The law's norms and values ultimately determine what models are appropriate, what questions should be asked of data, how burdens should be allocated, and what various statistical results imply.

Id. at 1113. That is a caution that everyone should heed.

cated. As Richard Epstein has pointed out, although we abhor murder—indeed, we deem it the greatest moral transgression that can occur—we do not respond to that abhorrence by relaxing the evidentiary rules governing murder trials despite the possibility that ordinary rules of proof may cause some murderers to go free.⁷⁷ In the discrimination context, however, we constantly hear arguments that evidentiary requirements or substantive rules should be eased in order to avoid the possibility that some clever defendants are discriminating and covering their tracks.⁷⁸ Unless we are willing squarely, and publicly, to embrace a policy preference for quota hiring—a course that seems politically unlikely⁷⁹—we should not covertly impose such a requirement under the pretense of formulating evidentiary rules.

In sum, contrary to the common assumption of courts to the contrary, statistical evidence of discrimination is inherently weak. Hypothesis testing is fundamentally incapable of doing that for which it is generally used—assessing the likelihood that disparities in an employer's work force were produced by nonchance factors. Acceptance by the courts of the Statistical Fallacy has caused them to assume away what in many cases will be the actual explanation for a statistical disparity—the operation of chance. Employers are thus left in the unenviable position of having to demonstrate a systematic nondiscrim-

77. As Richard Epstein has commented: "Why should the (assumed) importance of the antidiscrimination laws require us to slight the errors of overenforcement? The consensus that murder is a grave wrong, punishable under the criminal laws, has never been regarded as a reason to make life easy for prosecutors . . ." RICHARD A. EPSTEIN, *FORBIDDEN GROUNDS: THE CASE AGAINST EMPLOYMENT DISCRIMINATION LAWS* 225 (1992).

78. The distinction between treatment of potentially vindicated murder defendants and discrimination defendants is not simply a difference in philosophy between civil and criminal standards of proof. In the context of rape we hear arguments similar to those advanced in the discrimination context—that elements of the crime should be presumed or modified because convictions are otherwise too difficult to obtain and that evidentiary rules should be altered to the detriment of the defendant. Cf. Susan Estrich, *Rape*, 95 YALE L.J. 1037, 1104 (1986) (arguing for a standard under which a woman's "no means no," with no further inquiry into the issue of consent permissible); *State ex rel. M.T.S.*, 609 A.2d 1266, 1279 (N.J. 1992) (going beyond Estrich by holding, in effect, that absence of affirmative consent means "no": "the law places no burden on the alleged victim to have expressed nonconsent or to have denied permission, and no inquiry is made into what he or she thought or desired or why he or she did not resist or protest"); Frank Tuerkheimer, *A Reassessment and Redefinition of Rape Shield Laws*, 50 OHIO ST. L.J. 1245-54 (1989) (arguing that many rape shield laws exclude relevant evidence). Thus, rather than being a distinction between civil and criminal rules, the critical issue for those who would modify the substantive and procedural rules appears to be whether their conception of equality is offended by the specter of falsely exculpated defendants.

79. See Kingsley R. Browne, *The Civil Rights Act of 1991: A "Quota Bill," a Codification of Griggs, a Partial Return to Wards Cove, or All of the Above?*, 43 CASE W. RES. L. REV. 287 (1992).

inatory reason for work-force disparities when no systematic explanation exists.

III. THE CENTRAL ASSUMPTION

In the prior section, we were largely concerned with disparities in the roughly two to three standard deviation range—with associated probabilities of roughly 5 percent to less than 1 percent—which are present in a very large number of discrimination cases.⁸⁰ Random disparities of this magnitude are pervasive in the workplace and are not suggestive of a nonrandom cause, let alone an illegal one. However, in many cases work-force disparities are much greater, perhaps in the one in one million range.⁸¹ Although this Article's central thesis still holds—i.e., the statistical showing does not demonstrate that when a disparity of this size is observed there is only a one in a million chance that it occurred by chance—nonetheless one might reasonably conclude that chance is an unlikely explanation even if a precise probability cannot be assigned to a nonrandom cause. Although chance cannot be ruled out completely as the cause, a systematic explanation may be more plausible. It is in cases such as this that the second fundamental assumption of statistical proof becomes so critical.

The factual assumption that a non-discriminating employer's work force would (except for chance variations) mirror the race, ethnic, sex, and age profile of the "qualified" population—what I call the Central Assumption—is an essential foundation of a theory that allows an inference of discrimination to be drawn from statistical disparities.⁸² Tests of statistical significance are deemed to exclude chance as a cause of the disparities, and the assumption that populations are equally qualified and interested in each job excludes differential qualifi-

80. See, e.g., *Mozee v. American Commercial Marine Serv. Co.*, 940 F.2d 1036, 1043 (7th Cir. 1991) (promotion statistics reveal a 2.13 standard deviation disparity), *cert. denied*, 113 S. Ct. 207 (1992); *Rendon v. AT&T Technologies*, 883 F.2d 388, 397 (5th Cir. 1989) (rejecting argument that there is a strict benchmark of 3 standard deviations). In several discrimination cases, the statistical significance has been so low that the courts were forced to decide whether a one-tailed or two-tailed test of significance was appropriate. See, e.g., *Palmer v. Shultz*, 815 F.2d 84, 95 (D.C. Cir. 1987) (opting for a two-tailed test).

81. For example, in *EEOC v. Sears, Roebuck & Co.*, 839 F.2d 302, 323 n.20 (7th Cir. 1988), the z-values for nationwide comparisons ranged from 11.9 to 45.1.

82. See *BALDUS & COLE*, *supra* note 39, § 9.02 ("[I]n order to assess the likelihood that chance was the causal factor in a disparate treatment case, it is necessary to assume that all applicants were similarly situated on all relevant qualifications.").

cations and interest as explanations for the disparities. Courts are then left with discrimination as the only possible explanation.⁸³

In cases where the statistical probability of a given result is in the 1-to-5-percent range, the statistical evidence is so weak that the erroneous factual assumption is virtually irrelevant. Put another way, if I am right about the Statistical Fallacy but wrong in my criticism of the Central Assumption, disparities of two to three standard deviations are very weak evidence of discrimination, too weak to constitute a *prima facie* case. In the "megadisparity" cases, however, the fact that one cannot establish a precise probability that a disparity was caused by nonrandom factors may not substantially diminish the inference that the observed disparity was caused by something other than chance. To take an extreme example, if the probability of a result given random selection is one in one billion, it is quite likely that we would not observe the result in our lifetime if selection is purely random. In light of what we know of the prevalence nonrandom causes—such as differential qualifications, differential interest, and discrimination—we may subjectively, and correctly, believe that a nonrandom explanation is more likely than a random one. The effect of the Central Assumption, however, is to presume that once chance is excluded as an explanation, the nonrandom cause is an illegal one, thereby, in effect, imposing the burden on the employer to disprove discrimination.

As a threshold matter, it should be noted that the Central Assumption of fungibility of workers is directly at odds with other assumptions of discrimination law, specifically those dealing with affirmative action and the disparate-impact theory of discrimination. This inconsistency does not, of course, demonstrate that the assumption is wrong, but it might cause one to wonder why different employment discrimination doctrines have developed based upon inconsistent factual assumptions. The core problem with the Central Assumption, however, is that it is demonstrably wrong. Members of various racial, ethnic, sex, and age classifications are not equally qualified and inter-

83. It is important to emphasize that as soon as the argument turns from whether a disparity was caused by chance to what nonchance factor was responsible—i.e., a discriminatory or nondiscriminatory factor—the level of statistical significance is no longer particularly important. That is, if the plaintiff argues that a given disparity was caused by discrimination and the defendant argues that it was caused by differential interest, whether the disparity is three standard deviations or thirty does not affect the likelihood that either the plaintiff or defendant is correct. Both parties agree that the disparity is not caused by chance; that is, they agree that there really is a difference between the two groups.

ested in each job.⁸⁴ Especially at a time when we are constantly told of the virtues of “diversity”—i.e., that persons of different groups have different perspectives and attitudes—one would think that it would be similarly recognized that they may have different interests and abilities as well. At the least, it should be recognized that there is no reason *a priori* to structure our legal system with the presumption that they are identical.

A. *The Central Assumption and Other Title VII Doctrine*

The Central Assumption is fundamentally inconsistent with assumptions underlying other Title VII doctrine. For example, advocates of affirmative action have long argued that simple compliance with the nondiscrimination mandate of Title VII will not achieve full integration of women and minorities into the work force in the foreseeable future.⁸⁵ Because historically disfavored groups—especially blacks—are so disadvantaged by deficits in education and employment, the argument goes, without affirmative action full integration will take generations. Thus, the rationale underlying affirmative action is exactly contrary to the rationale supporting statistical inferences of discrimination. The former posits that proportional representation is not, at least in the near term, the expected consequence of nondiscrimination; the latter posits that it is.

The Central Assumption is also contrary to the rationale of the disparate-impact theory of discrimination. Under disparate-impact theory, a plaintiff may challenge practices having a disproportionate effect on a particular group even if the practices were not adopted for the purpose of discrimination.⁸⁶ In *Griggs v. Duke Power Co.*,⁸⁷ the Supreme Court stated that employment qualifications such as satisfactory scores on standardized tests and possession of a high school diploma may “operate as ‘built-in headwinds’ for minority groups.”⁸⁸ A central tenet of Title VII doctrine, as it has been developed by the courts, is that qualifications adopted without an intent to exclude may nonetheless have the effect of excluding minorities and women. As a result, employers may not adopt qualifications causing a disparate

84. See Smith & Abram, *supra* note 13, at 38 (“Many of the observed shortfalls in minority and female earnings have been found to be attributable to differences in supply side characteristics among minorities, women, and white males . . . [and] result from forces beyond the control of an individual employer.”).

85. See generally Randall Kennedy, *Persuasion and Distrust: A Comment on the Affirmative Action Debate*, 99 HARV. L. REV. 1327 (1986).

86. See *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971).

87. 401 U.S. 424 (1971).

88. *Id.* at 432.

impact unless the qualifications are justified by business necessity. Yet, if the Central Assumption were correct, there would be no need for disparate-impact theory, since qualifications, even unnecessary ones, would be evenly distributed throughout all groups, and the only disparate impacts would be random ones, as likely to favor blacks as whites. However, no one believes that the effects are random. Numerous disparate-impact cases have been brought alleging that qualifications adopted by employers without discriminatory intent have resulted in disproportionate exclusion. In some cases, employers are able to defend on the ground of business necessity; in others they are not. The relevant point, however, is that many job qualifications adopted without discriminatory intent will have a disparate impact on certain groups. Thus, the disparate-impact rationale is also directly contrary to the Central Assumption, with the former positing that good-faith reliance on qualifications will tend to result in a disproportionately low rate of selection of women and minorities, and the latter positing that it will not.

B. The Factual Basis for the Central Assumption

The Central Assumption of equivalence in its rawest form has few adherents. That is, no one believes that a minority plucked from the ghetto of the inner city is as likely as a white from the affluent suburbs to be qualified for a position as a lawyer. For that reason, except in cases involving jobs requiring no skills (or requiring only readily obtainable skills), raw comparisons between an employer's work force and the general population are generally recognized to be inappropriate.⁸⁹ On this gross level, then, no one really believes the *Teamsters* statement that "nondiscriminatory hiring practices will in time result in a work force more or less representative of the racial and ethnic composition of the population in the community from which employees are hired," unless one gives extraordinary weight to the phrases "in time" and "more or less." Instead, at least in the absence of applicant-flow data, courts will generally require comparisons between the

89. Courts that have relieved plaintiffs of the obligation to prove qualifications in circumstances where job skills are readily obtainable have not adequately explained why the fact that skills are readily obtainable justifies declining to control for them. *See, e.g., Gay v. Waiters' & Dairy Lunchmen's Union, Local No. 30*, 489 F. Supp. 282, 307 (N.D. Cal. 1980) (holding that because the necessary skills of a waiter "can be acquired on the job in a matter of days or weeks . . . general population and civilian work force data are the proper frame of reference"), *aff'd*, 694 F.2d 531 (9th Cir. 1982). The fact that employers require employees to have the skills for the job, rather than providing on-the-job training, surely is not suggestive of intentional discrimination.

employer's work force and the "qualified labor force."⁹⁰ The statistical analysis of intentional discrimination then assumes that within this "qualified labor force" qualifications and interest are randomly distributed with respect to race and sex. Limiting the statistical comparison to this qualified labor force is thought to control for differences in qualifications between groups.

The assumption that qualifications are randomly distributed by race and sex within the qualified labor force has no more empirical basis than the same assumption with respect to the general population. The qualifications that define the qualified labor force are so generalized as to belie any claim that there is serious control for qualifications. For example, if the employer requires a law degree for a position, then the qualified labor force would be those persons having such a degree, the ostensible assumption being that qualifications are now controlled for.⁹¹ Yet, how many of the law professors who argue for holding employers liable on the basis of such statistics think that all law school graduates are equally qualified to be hired on their faculties? How many believe that the qualifications of a randomly selected graduate of the "worst" law school are likely to be equal to those of a randomly selected graduate of the "best" law school? How many believe that the person who graduates at the top of his class from a given law school is no better than the person who graduates at the bottom of the same class? How many believe—*notwithstanding* discrimination in education, differences in attitudes toward education, and any other reasons that may cause racial differences in educational achievement—that blacks and whites on average will have equal law school credentials? The tendency of law schools to hire highly credentialed applicants and to hire "affirmative-action candidates" with lesser

90. In *EEOC v. Chicago Miniature Lamp Works*, 947 F.2d 292 (7th Cir. 1991), the district court had found that the defendant's applicant flow had to be ignored because it was "not remotely representative of the racial and ethnic composition of the civilian work force." *Id.* at 300. Of course, whenever it is argued, by either plaintiff or defendant, that applicant-flow data rather than census data should be employed, it will be precisely because the applicant flow differs from the census data.

Issues concerning the appropriate geographic scope of the relevant labor market are beyond the scope of this Article, but definition of that market is often outcome determinative. *Id.* at 299–300; *see also* Braun, *supra* note 17, at 63–67 (questioning utility of EEOC's use of the "Standard Metropolitan Statistical Area" as the proper geographic area to be considered); Smith & Abram, *supra* note 13, at 59 ("Egregious displays of gerrymandering characterize litigation and compliance proceedings.").

91. *See, e.g.,* *Valentino v. United States Postal Serv.*, 674 F.2d 56, 71 n.24 (D.C. Cir. 1982) (stating that plaintiff need control only for "threshold qualifications," which are "those prerequisite to consideration for a position (e.g., a law degree), not attributes of relative advantage or disadvantage (e.g., reputation of law school attended, service on a law journal)").

credentials suggests that most faculty members believe that credentials are important and not uniformly distributed throughout the population. Yet the core principle of the Central Assumption is that those with the minimum qualifications are essentially fungible and that reliance on higher qualifications is likely to have a race- and sex-neutral result.

Some would acknowledge that law school graduates are not fungible but assume that applicants for other positions are⁹²—an attitude that allows them to impose requirements on others that they correctly perceive would be inappropriate for themselves. However, the notion that variations in qualifications for “lower level” jobs are not important is born of ignorance and elitism—an attitude that “they” can live under rules that would be inappropriate for “us,” because we are more important and because there are differences in our group that do not exist among the “drones.” However, there are good janitors and bad janitors, just as there are good lawyers and bad lawyers. An employer’s interest in choosing between good and bad janitors is as legitimate as its interest in choosing between good and bad lawyers.⁹³

The assumption that race and ethnicity are unrelated to productivity is so demonstrably false as to require little refutation.⁹⁴ For example, we are constantly told that blacks attend inferior schools, tend to get less education, do less well on standardized achievement tests and, one suspects, although data are not as readily available, do less well in schools even when they finish.⁹⁵ The last conclusion would not be surprising given that blacks are often admitted to institutions of higher education with lower grades and test scores, whether or not the insti-

92. See Elaine W. Shoben, *Probing the Discriminatory Effects of Employee Selection Procedures with Disparate Impact Analysis Under Title VII*, 56 TEX. L. REV. 1, 33 (1977) (suggesting that “[p]olice officers, fire fighters, many factory workers, and bank tellers” have jobs requiring “skills commonly possessed or easily attained by many people”).

93. It is true, of course, that for more complex jobs, an appropriate statistical comparison would contain more variables than an analysis of less complex jobs. See *Coser v. Moore*, 739 F.2d 746, 750 (2d Cir. 1984) (“[A] university’s ‘treatment of minorities cannot be evaluated in the same manner as that of, say, a trucking company or factory that can hire many people with less differentiated qualifications.’”) (quoting the district court opinion, 587 F. Supp. 572, 584 (E.D.N.Y. 1983)). That is not the same as saying that applicants for the lower level jobs are fungible.

94. See DANIEL S. HAMERMESH & ALBERT REES, *THE ECONOMICS OF WORK AND PAY* 360–71 (1988) (discussing differences in productivity of workers in different racial, ethnic, and sex groups); THOMAS SOWELL, *THE ECONOMICS AND POLITICS OF RACE: AN INTERNATIONAL PERSPECTIVE* 248–50 (1983).

95. See generally ANDREW HACKER, *TWO NATIONS: BLACK AND WHITE, SEPARATE, HOSTILE, UNEQUAL* (1992); James P. Smith & Finis R. Welch, *Black Economic Progress After Myrdal*, 27 J. ECON. LITERATURE 519 (1989).

tutions are willing to admit it.⁹⁶ With respect to job experience, we hear often of the high level of unemployment and underemployment of blacks, which has a substantial impact upon their predicted productivity to a prospective employer.⁹⁷

In addition to differences in productivity-related qualifications, there is reason to think that there may also be racial and ethnic differences in interest in various jobs; at least there is no basis to erect a legal presumption that such differences do not exist. For example, in light of the often-commented-upon suspicion that blacks have of police, one would not necessarily expect equal levels of interest in joining the police force. Conversely, given the strong tradition of police service in many Irish communities,⁹⁸ the fact that they are represented at a rate substantially higher than their rate of participation in the general labor force should raise no suspicion. Similarly, Jews are found in high numbers in law teaching,⁹⁹ a fact that does not lead most to suspect discrimination against gentiles.

Just as with race, productivity-related differences exist between the sexes.¹⁰⁰ There are differences in education; for example, men earn the great majority of Ph.D.'s in engineering, physics and mathematics,¹⁰¹ and in the absence of affirmative action in graduate admissions the sex

96. See, e.g., Michel Marriott, *Unresolved: Role of Race in Law Class Admissions*, N.Y. TIMES, April 28, 1991, § 4, at 5 (reporting statement by the dean of Georgetown Law School that race is not among the factors considered for admission). See generally Lino A. Graglia, *Race Norming in Law School Admissions*, 42 J. LEGAL EDUC. 97 (1992).

97. In *Coates v. Johnson & Johnson*, 756 F.2d 524 (7th Cir. 1985), the question was whether statistics showing a disproportionately high rate of disciplinary sanctions imposed on blacks suggested discrimination. However, it is not clear that we should suspect discrimination in workplace discipline simply because more blacks are disciplined than whites. Most people do not generally assume—although some would argue to the contrary—that the criminal justice system is infected with racism merely because blacks are disproportionately represented in the ranks of convicted violent criminals; the more likely explanation is that blacks are disproportionately represented in the ranks of *all* violent criminals. If a group that contributes disproportionately to violent crime is also found to be punished disproportionately for workplace misconduct, it is not immediately obvious that discrimination by the disciplining employer is the most likely explanation. Yet, the statistical method of testing for discrimination assumes that it is.

98. GEORGE E. REEDY, *FROM THE WARD TO THE WHITE HOUSE: THE IRISH IN AMERICAN POLITICS* 28 (1991).

99. ROBERT A. BURT, *TWO JEWISH JUSTICES: OUTCASTS IN THE PROMISED LAND* 64 (1988) (noting that in 1970, 25% of law teachers nationwide were Jews and that at "elite" law schools the figure was 38%).

100. June O'Neill, *The Trend in the Male-Female Wage Gap in the United States*, J. LAB. ECON., Jan. 1985 Supp., at S91, S92 (1985) (arguing that the wage gap between men and women "did not narrow between 1955 and 1982 because the average sex differential in productivity-related characteristics did not, on balance, narrow").

101. In 1989, for example, women received 18.1% of the Ph.D.'s in mathematics, 18.9% in physical sciences, and 8.2% in engineering. STATISTICAL ABSTRACT OF THE UNITED STATES 597 (1991).

imbalance would be even greater. There are also differences in employment experience; for example, the disproportionate responsibility of women for household work and child care causes many women to devote less effort to market work than men and to evaluate certain job characteristics differently, which, in turn, has an impact on their occupational distributions and their earnings.¹⁰²

The assumption of equal interest in all jobs between men and women also has no basis in fact and is counter to everyday experience.¹⁰³ One need only look around the contemporary workplace to observe a substantial degree of occupational segregation, most of which cannot be attributed to employer discrimination. As Deborah Rhode has pointed out, by early adolescence boys and girls have acquired different career expectations.¹⁰⁴ Sex differences in temperament, abilities, and worldview are well documented¹⁰⁵ and are commonly invoked by feminists in support of their social analyses.¹⁰⁶

102. O'Neill, *supra* note 101, at S110-11; *see also* HAMERMESH & REES, *supra* note 94, at 368 ("The interruption of careers by child rearing means that at any given age married women are likely to have had less labor-force experience than men, and hence to have made smaller investments in on-the-job training.").

103. *See generally* Randall K. Filer, *Sexual Differences in Earnings: The Role of Individual Personalities and Tastes*, 18 J. HUM. RESOURCES 82 (1983).

104. DEBORAH L. RHODE, *JUSTICE AND GENDER: SEX DISCRIMINATION AND THE LAW* 166 (1989). She states: "In general, women have expressed lower expectations for occupational success than men and have attached greater priority to the relational aspects of employment (such as opportunities for helping or working with others) than to opportunities for recognition (money, status, and power)." *Id.*

105. *See generally* CAROL GILLIGAN, *IN A DIFFERENT VOICE: PSYCHOLOGICAL THEORY AND WOMEN'S DEVELOPMENT* (1982) [hereinafter *IN A DIFFERENT VOICE*]; Kingsley R. Browne, *Biology, Equality, and the Law: The Legal Significance of Biological Sex Differences*, 38 SW. L.J. 617 (1984).

106. For example, Carol Gilligan's work, *IN A DIFFERENT VOICE*, *supra* note 105, is commonly cited in the legal literature for the proposition that men and women have different values and modes of thought. *See, e.g.*, Robin West, *Jurisprudence and Gender*, 55 U. CHI. L. REV. 1 (1988):

[According to "cultural feminists,"] women have a "sense" of existential "connection" to other human life which men do not. That sense of connection in turn entails a way of learning, a path of moral development, an aesthetic sense, and a view of the world and of one's place within it which sharply contrasts with men's.

Id. at 15; Ann E. Freedman, *Sex Equality, Sex Differences, and the Supreme Court*, 92 YALE L.J. 913 (1983):

Another component of a new sex discrimination jurisprudence must be an awareness that assimilation into existing predominantly male social structures is an inadequate definition of equality between the sexes and one that robs equality of much of its transformative potential. The principle that only women whose life patterns, skills, and experiences are virtually identical to those typical of men will be accorded high status and rewards will, as a practical matter, doom most women to continued subordination. Moreover, the common practice of focusing primarily on fair access to previously male positions and privileges is based in large part on androcentric value systems that maintain the hierarchy of male over female activi-

Although we are repeatedly told that women are differently socialized,¹⁰⁷ differently educated, and even that they think differently, the statistical analysis employed in discrimination cases erects a strong presumption that without discrimination by employers, women's career paths would be identical to those of men. Yet if these differences are real, it should hardly be surprising if the career choices of men and women are not entirely congruent;¹⁰⁸ indeed, it would be surprising if they were. The point here is not, as Vicki Schultz has derisively characterized it,¹⁰⁹ that "because girls are conditioned to conform to 'feminine' sex roles, adult women will automatically aspire to 'feminine' work." Neither women nor men "automatically aspire" to anything. The point is considerably more modest; it is that recognition of these differences should cause one to be skeptical of a legal presumption that women will "automatically aspire" to the same work as men.¹¹⁰ Whether these differences in preferences result from innate predisposition¹¹¹ or from "cultural stereotypes, family and peer pressure, and the absence of alternative role models,"¹¹² there is nothing in

ties. Since women's value systems are often quite different from those of men, the full participation of women in the process of change may both depend on and encourage a re-ordering of social priorities.

Id. at 967.

107. See RHODE, *supra* note 104, at 165–66 ("At early ages, children begin absorbing cues about appropriate sex-role traits and occupations . . .").

108. *Id.* at 164 ("To assume that under conditions of full equality women will make precisely the same occupational choices as men is to adopt an assimilationist perspective that many feminists renounce.").

109. Vicki Schultz, *Telling Stories About Women and Work: Judicial Interpretations of Sex Segregation in the Workplace in Title VII Cases Raising the Lack of Interest Argument*, 103 HARV. L. REV. 1749, 1810 (1990).

110. Schultz has also argued that courts have erroneously attributed women's work aspirations to their early socialization and have failed to understand that women's work aspirations are shaped in large part by later experiences in the labor market. *Id.* at 1799. That fact, if it is one, is largely irrelevant to her point. Even if women have less interest in a particular kind of job because of their labor market experience, differential interest should not be ignored in a statistical analysis of a particular employer's work force, where the question is not whether women have faced discriminatory attitudes in the labor market but whether they have faced discrimination *from this employer*. Ignoring differential interest because of "sexist societal attitudes" is another way of modifying the substantive reach of Title VII under the guise of creating evidentiary rules. Under formal Title VII doctrine, employers are liable for their own discrimination, not for that of others. Yet excluding differential interest from a statistical analysis has the effect of holding employers liable for actions of others. Schultz acknowledges that part of her goal in urging courts to reject defenses based upon differential interest is to pressure employers to recruit women into nontraditional jobs. However, Title VII was intended to *allow* women to choose nontraditional jobs, not to require them to do so or to require employers to attempt to manipulate women's choices.

111. See generally Browne, *supra* note 79.

112. RHODE, *supra* note 104, at 166.

the language or history of Title VII suggesting that employers should be held responsible for these differences.

Courts have employed the presumption of equal interest in circumstances where it is clearly contrary to reality. For example, in *Catlett v. Missouri Highway and Transportation Commission*,¹¹³ the Eighth Circuit held that the relevant comparative pool against which to compare Missouri's hiring for highway maintenance workers was the general labor force. As a result, the fact that less than ten percent of the state's maintenance hires were women, compared with forty-eight percent of the labor force, was held to suggest discrimination.¹¹⁴ Yet, the notion that in the absence of discrimination women would flock to such jobs in proportion to their representation in the total labor force is nonsensical.¹¹⁵

Productivity-related differences also exist for persons at different ages. Decline in certain kinds of productivity with age is well documented.¹¹⁶ As a result, it should not be surprising that when employers make reduction-in-force decisions based upon performance, often a fairly significant number of older workers are targeted for layoff. Moreover, for any given job containing workers of varying ages, some of the employees will have peaked in that particular job, while others ("the rising stars") will just be passing through on their way to better things. A sixty-year-old worker who has been in a middle-management position for fifteen years is likely to be a less valuable employee than the thirty-year-old who was recently promoted into that position

113. 828 F.2d 1260, 1266 (8th Cir. 1987), *cert. denied*, 485 U.S. 1021 (1988).

114. *Id.*

115. See *Johnson v. Transportation Agency*, 480 U.S. 616, 668 (1987) (Scalia, J., dissenting) ("It is absurd to think that the nationwide failure of road maintenance crews . . . to achieve the Agency's ambition of 36.4% female representation is attributable primarily, if even substantially, to systematic exclusion of women eager to shoulder pick and shovel."); *Dothard v. Rawlinson*, 433 U.S. 321, 348 (1977) (White, J., dissenting) ("I am unwilling to believe that the percentage of women applying or interested in applying for jobs as prison guards in Alabama approximates the percentage of women either in the national or state population.").

The district court in *Davis v. City of Dallas*, 483 F. Supp. 54 (N.D. Tex. 1979), used a more realistic approach than the *Catlett* court by relying on applicant-flow data, rather than labor force statistics, in reviewing a claim of sex discrimination in the Dallas Police Department:

The most plausible explanation for the disparity, an explanation borne out by the evidence discussed, is that women were not during the relevant period interested in police work in the same proportion as were men. . . .

These job preferences of females may be born of attitudes conditioned by societal sexist values. But frustration with the realization that equality of opportunity untouched by gender remains a social goal and not an achieved reality, must not be visited on this employer in the form of liability.

Id. at 61.

116. MARY JABLONSKI ET AL., PRODUCTIVITY, AGE, AND LABOR COMPOSITION CHANGES IN THE U.S., 111 Monthly Lab. Rev. No. 9, at 34 (1988).

and who may be perceived as having the potential to rise much higher in the organization.¹¹⁷ Nonetheless, with the kind of crude statistical comparisons involved in pattern-or-practice cases, these two employees are likely to be considered equal,¹¹⁸ or, even worse, the older employee may be viewed as *more* qualified if seniority or experience is considered a positive qualification.

In sum, the Central Assumption of equal interest and qualifications between different race, sex, and age groupings is contrary to both empirical evidence and intuition. Yet it is enshrined in our case law and serves as the basis for inappropriately transferring to employers the burden of proving its falsity in individual cases.

C. The Central Assumption and Improper Shifting of the Burdens of Proof

Permitting plaintiffs to establish a prima facie case based upon incomplete statistics in effect shifts the burden of proof to the defendant to disprove matters that should be part of the plaintiffs' case in chief.¹¹⁹ In a case of individual disparate treatment under *McDonnell Douglas Corp. v. Green*,¹²⁰ a plaintiff, in order to establish a prima facie case, is generally obligated to show that he was qualified for the job, that he applied for it, and that he was denied it under circumstances that raise an inference of discrimination, this last step generally entailing a showing that the plaintiff was at least as qualified as the person who obtained the job (typically a white or a male).¹²¹ It is not

117. See *CONNOLLY ET AL.*, *supra* note 9, § 10.05[1], at 10-37 ("[T]o the extent that age and seniority are correlated, it is . . . not unreasonable to expect that within a job level, the older employees will tend to be the less able performers."); see also *Davidson v. Board of Governors*, 920 F.2d 441, 446 (7th Cir. 1990) ("Employees who stay longest are, on the average though of course not in every case, those who have the fewest offers to go elsewhere.").

118. See *Barnes v. GenCorp, Inc.*, 896 F.2d 1457, 1466 (6th Cir.) ("Since it is reasonable to presume at [the prima facie case] stage of the case that skill is distributed randomly over any given age group and since the plaintiffs have shown that the results depart significantly from those that chance alone would predict, we believe that the statistics, on their face, establish a prima facie case of discrimination."), *cert. denied*, 111 S. Ct. 211 (1990).

119. See *Campbell*, *supra* note 41, at 1309 ("To shift the burden of persuasion to the defendant . . . on the strength of a poorly fitted regression violates the burden of proof standards enunciated in *Texas Department of Community Affairs v. Burdine* and *Furnco [Constr. Co. v. Waters]*, even if the defendant has offered no better fitting equation.").

120. 411 U.S. 792 (1973).

121. The precise contours of the plaintiff's prima facie case differ depending upon the nature of the hiring decision. Under the facts of *McDonnell Douglas* itself, the Court held that to establish a prima facie case, the plaintiff must show:

(i) that he belongs to a racial minority; (ii) that he applied and was qualified for a job for which the employer was seeking applicants; (iii) that, despite his qualifications, he was rejected; and (iv) that, after his rejection, the position remained open and the employer continued to seek applicants from persons of complainant's qualifications.

enough for the plaintiff merely to demonstrate that he is a minority, that he met the minimum qualifications, and that he did not get the job. Yet, that is exactly what courts permit when they allow the typical statistical showing to satisfy the prima facie standard.¹²²

The oft-repeated assertion that the plaintiffs' burden in a pattern-or-practice case is greater than the plaintiff's burden in an individual case¹²³—an assertion generally made for the purpose of imposing a

Id. at 802.

In circumstances where the employer reviews applications all at one time and selects from the applications, the mere rejection of a minimally qualified minority should not be enough to establish a prima facie case, since rejection of one qualified applicant in favor of another raises no inference of discrimination. In such circumstances, courts often require as part of the plaintiff's showing that the plaintiff either demonstrate that he was at least as qualified as the nonminority selected or provide other circumstantial evidence of discrimination. *See, e.g., Hill v. Seaboard Coast Line R.R. Co.*, 885 F.2d 804, 809 (11th Cir. 1989) (holding that plaintiff must show that he is at least as qualified as the nonminority selected); *Mason v. Continental Ill. Nat'l Bank*, 704 F.2d 361, 365 (7th Cir. 1983) (suggesting in dictum that "there is some question whether the [*McDonnell Douglas*] format is appropriate . . . in a case involving appointment or promotion to a position for which there are several candidates"); *Loeb v. Textron, Inc.*, 600 F.2d 1003 (1st Cir. 1979):

[I]f a qualified black were turned down for a job for which there were 100 applicants for every opening and the practice, rather than being to hire the first qualified applicant who appeared, was to choose on the basis of qualifications, recommendations and subjective impressions gleaned from an interview, it would hardly seem 'more likely than not' that the applicant was rejected because of race.

Id. at 1017 n.18; *Bracey v. Helene Curtis, Inc.*, 780 F. Supp. 568, 571 (N.D. Ill. 1992) ("An inference of racial discrimination cannot arise from the mere fact that a white person was hired instead of [plaintiff] from among a pool of applicants."); *Shidaker v. Bolger*, 593 F. Supp. 823, 838 (N.D. Ill. 1984) ("[O]utside the factory setting the first qualified applicant need not be selected where a choice is made 'on the basis of qualifications, recommendations and subjective impressions.'") (quoting *Mason*, 704 F.2d at 365), *aff'd in part and rev'd in part sub nom. Shidaker v. Carlin*, 782 F.2d 746 (7th Cir. 1986), *vacated sub nom. Tisch v. Shidaker*, 481 U.S. 1001 (1987)). *But see Garner v. Boorstin*, 690 F.2d 1034, 1036 n.4 (D.C. Cir. 1982) (In a promotion case, a black plaintiff established a prima facie case by showing that he was at least minimally qualified and that "the available positions were filled by individuals with comparable qualifications who were not members of the classes protected by the relevant statutes."); *Davis v. Weidner*, 596 F.2d 726, 730 (7th Cir. 1979) (holding that even where the employer's decision involves a "simultaneous choice between prospective employees on the basis of relative qualifications," rejection of woman establishes a prima facie case).

122. *See Palmer v. Shultz*, 815 F.2d 84, 91 n.6 (D.C. Cir. 1987) ("[A]s long as a plaintiff's statistical analysis has properly defined the pool of eligible candidates, by accounting for 'minimum objective qualifications,' the burden then shifts to the defendant to introduce evidence of a legitimate, nondiscriminatory explanation if the analysis reveals a statistically significant disparity.").

123. *See, e.g., Segar v. Smith*, 738 F.2d 1249, 1302 (D.C. Cir. 1984) ("The initial proof demanded of plaintiffs in a pattern and practice case typically goes far beyond the minimal showing that is required for an individual to establish a prima facie case under *McDonnell Douglas*."), *cert. denied sub nom. Meese v. Segar*, 471 U.S. 1115 (1985). In *Craik v. Minnesota State University Board*, 731 F.2d 465, 471 n.9 (8th Cir. 1984), the court stated that "[t]he prima facie case established by a finding that an employer is guilty of a pattern or practice of discrimination goes far beyond the prima facie case contemplated by *McDonnell Douglas* and

heightened burden of rebuttal on defendants¹²⁴—is simply not true; in fact, in many ways the truth is just the opposite. The pattern-or-practice prima facie case *assumes* most of the elements of the *McDonnell Douglas* prima facie case, and it assumes them on a grand scale. When a plaintiff compares the employer's work force to "qualified labor market" statistics, the analysis in effect eliminates the need to show that the plaintiffs applied for the job; instead it is presumed that members of all groups applied at the same rate.¹²⁵ Second, the structure of the prima facie case eliminates the need to show that the plaintiffs were as, or more, qualified than those who obtained the jobs; instead, the analysis presumes that black applicants and white applicants are equally qualified, even though at best the analysis has controlled only for minimum qualifications or, even worse, for proxies for minimum qualifications.¹²⁶

Burdine." The court suggested that the Supreme Court in *International Brotherhood of Teamsters*, 431 U.S. 324 (1977), had recognized the strength of the plaintiffs' prima facie case: "[a] finding of a discriminatory pattern or practice creates 'a greater likelihood that any single decision was a component of the overall pattern,' and changes 'the position of the employer to that of a proved wrongdoer'" *Craik*, 731 F.2d at 471 n.9 (quoting *Teamsters*, 431 U.S. at 359–60 n.45).

The *Craik* court misunderstood the difference between a *Teamsters* prima facie case and the ultimate finding of a pattern or practice of discrimination. The Supreme Court in *Teamsters* had said that *proof* that discrimination is the employer's standard operating procedure makes it likely that each decision of the employer is a product of that procedure; therefore, at the remedy phase of a pattern-or-practice case, each class member is presumed entitled to a remedy, with the employer bearing the burden of persuasion to demonstrate nondiscrimination in a given decision. *Teamsters* did not suggest, as the *Craik* court implies, that the establishment of a prima facie case converts the status of the employer "to that of a proved wrongdoer."

124. See *Segar*, 738 F.2d at 1269 ("The bare articulation of a nondiscriminatory explanation, while sufficient to rebut an individual plaintiff's low-threshold *McDonnell-Douglas* showing, generally will not suffice as a rebuttal to a typical class-wide showing of pervasive discrimination" because the class "will typically have presented statistical evidence showing pervasive disparities and eliminating most, if not all, potential nondiscriminatory explanations for the observed disparities.").

125. Elaine W. Shoben, *In Defense of Disparate Impact Analysis Under Title VII: A Reply to Dr. Cohn*, 55 IND. L.J. 515, 519 (1980).

126. See *Segar*, 738 F.2d at 1299 ("[P]laintiffs need not control for specific minimum job qualifications, as long as their data include a reasonable proxy for these qualifications."); see also *Taylor v. Teletype Corp.*, 648 F.2d 1129, 1133 (8th Cir.) (using "layoff" as a proxy for "demotion"), *cert. denied*, 454 U.S. 969 (1981); *Vuyanich v. Republic Nat'l Bank*, 505 F. Supp. 224, 314–17 (N.D. Tex. 1980) (using "age" as a proxy for "prior work experience"), *vacated*, 723 F.2d 1195 (5th Cir.), *cert. denied*, 469 U.S. 1073 (1984).

One of the crudest form of proxies is reliance on census categories for determining the availability of qualified applicants. See *EEOC v. DuPont*, 445 F. Supp. 223 (D. Del. 1977):

The Government suggests that those which the census classifies as "professional, technical, and kindred workers" represent the segment of the labor force qualified to perform [defendant's] professional jobs. . . . [This category] includes dozens of groups whose background bears no relationship to the skills required by [defendant's] professional

Notwithstanding the elimination of the elements that make a *McDonnell Douglas* prima facie case at least weakly probative, belief in the power of statistical evidence has led many courts to impose a substantial burden of rebuttal on defendants in class cases. In a *McDonnell Douglas* case, the defendant faced with a prima facie case need only "articulate some legitimate, nondiscriminatory reason for the employee's rejection."¹²⁷ The defendant has no obligation to persuade the trier of fact that the articulated reason was the real reason; it is enough if its rebuttal "raises a genuine issue of fact as to whether it discriminated against the plaintiff."¹²⁸ In order to prevail, the plaintiff must then prove that the articulated reason is merely a pretext for discrimination.¹²⁹ The burden of persuasion remains on the plaintiff at all times. In contrast, many courts impose a substantially higher burden on defendants to rebut the prima facie case in a pattern-or-practice case. For example, because of its view that the prima facie case "show[s] pervasive disparities and eliminat[es] most, if not all, potential nondiscriminatory explanations for the observed disparities," the D.C. Circuit has held that "[t]he bare articulation of a nondiscriminatory explanation, while sufficient to rebut an individual plaintiff's low-threshold *McDonnell Douglas* showing, generally will not suffice as a rebuttal to a typical class-wide showing of pervasive discrimination."¹³⁰ Instead, the defendant must either "refute the plaintiffs' claim that a disparity exists" or explain that the "disparity has not resulted from illegal discrimination."¹³¹

Even the *McDonnell Douglas* prima facie case is only weakly probative of discrimination, but it is at least more probative than the statistical prima facie case. Although the *McDonnell Douglas* prima facie case is said to "rais[e] an inference of discrimination only because we presume these acts, if otherwise unexplained, are more likely than not

positions. . . . The data on professional, technical and kindred workers is, accordingly, inappropriate for use in evaluating [defendant's] record with respect to professional workers.

Id. at 246.

127. *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 802 (1973).

128. *Texas Dep't. of Community Affairs v. Burdine*, 450 U.S. 248, 254-55 (1981).

129. *Id.* at 255-56.

130. *Segar*, 738 F.2d at 1269; see also *Williams v. New Orleans S.S. Ass'n*, 673 F.2d 742, 750 n.14 (5th Cir. 1982) (Defendant's rebuttal burden in a pattern-or-practice case "is more onerous than the rebuttal burden required of defendants in individual Title VII claims under [*Burdine*]."), *cert. denied*, 460 U.S. 1038 (1983).

131. *Segar*, 738 F.2d at 1267-68. The *Segar* court acknowledged that "[t]he nature of the burden that the defendant bears on such a defense is not entirely free of doubt." *Id.* The same court in *MacKenzie v. Sawyer*, 684 F.2d 62, 71 n.7 (D.C. Cir. 1982), noted that "it remains debatable whether the *Burdine* holding that the burden of persuasion remains with an individual plaintiff should be applied to a class action suit alleging disparate treatment."

based on the consideration of impermissible factors,"¹³² no intelligent person could believe that the employer's failure to hire the first person through the door with the minimum qualifications is most likely a consequence of discrimination. Nonetheless, the fiction embodied in the *McDonnell Douglas* prima facie case is a useful and reasonably inexpensive one, because the employer's burden in rebuttal is relatively light; it is rare that the employer cannot even articulate a reason for its actions.¹³³ The fiction becomes much more expensive, however, when it results in an obligation on the employer's part to justify a statistical disparity involving perhaps hundreds or thousands of employees that may be a consequence of social forces unknown to the employer.¹³⁴ The employer cannot satisfy that burden simply by testifying that it has a practice of selecting the best applicant for the job,¹³⁵ and many courts will not allow employers to rebut the prima facie case merely by pointing to flaws in the plaintiffs' statistics.

Because productivity-related variables are often not distributed randomly with respect to race, sex, and age, the fewer such variables that are controlled for, the more likely it is that the statistical investigation will falsely yield positive results.¹³⁶ That is why qualified-labor-pool statistics are better than general-labor-force statistics, which in turn are better than general-population statistics.¹³⁷ However, as discussed below, even qualified-labor-pool statistics are virtually meaningless

132. *Furnco Constr. Corp. v. Waters*, 438 U.S. 567, 577 (1978).

133. In an individual case, once the defendant rebuts the prima facie case by articulating a legitimate nondiscriminatory reason for the action, the burden then shifts to the plaintiff to demonstrate that the defendant's justification is a "pretext." *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 804 (1973). It is on the pretext issue that most cases turn.

134. See *Smith & Abram*, *supra* note 13, at 47 ("This remarkable presumption—that evidence of work force imbalance is caused by unlawful discrimination, and is illegal unless justified—is at odds with our basic traditions.").

135. See, e.g., *International Bhd. of Teamsters v. United States*, 431 U.S. 324, 342–43 n.24 (1977); *Pegues v. Mississippi State Employment Serv.*, 699 F.2d 760, 766 (5th Cir.) ("Affirmations of good faith selection of the most qualified applicant will not satisfy [the employer's] burden."), *cert. denied*, 464 U.S. 991 (1983).

136. Failure to control for a variable that is correlated with protected status will falsely exaggerate the statistical disparity. See *Fisher*, *supra* note 41, at 708–09; James T. McKeown, *Statistics for Wage Discrimination Cases: Why the Statistical Models Used Cannot Prove or Disprove Sex Discrimination*, 67 IND. L.J. 633, 654 (1992); Norris, *supra* note 49, at 71 ("If the missing variable is differentially associated with a particular race or sex, its omission may incorrectly lead to a conclusion that discrimination exists."); see also *EEOC v. Sears, Roebuck & Co.*, 628 F. Supp. 1264, 1286 (N.D. Ill. 1986), *aff'd*, 839 F.2d 302 (7th Cir. 1988) ("[T]he assumptions made by a statistician in formulating a model can be far more important than the numerical complexities and results of the analysis."); Aickin, *supra* note 29, at 361 ("What is not contained in a regression equation can be very important—sometimes more important than the regression itself."). Conversely, however, controlling for too many variables can make it virtually impossible to achieve statistical significance unless the sample size is very large.

137. See *CONNOLLY ET AL.*, *supra* note 9, § 4.02[1].

because they fail to take into account the myriad of legitimate productivity-related variables and tend to focus instead on minimum qualifications. Many courts have been singularly inattentive to the consequences of permitting plaintiffs to establish a *prima facie* case based upon crude statistical showings and requiring the defendant to refine the showing. The remainder of this section will deal with ways in which courts have effectively shifted the burden of proof to employers and relieved plaintiffs of the obligation to introduce evidence that is rightfully a part of their own case.

1. *The Focus on Minimum Qualifications*

The most common formulation of the variables a plaintiff's analysis must consider is that the analysis must control for the "*minimum objective qualifications* for the positions at issue,"¹³⁸ a rule that goes far toward relieving the plaintiff of the obligation to introduce statistical analyses that can in any sense be said to be probative of discrimination. The justification for this rule, as described by the D.C. Circuit in *Segar v. Smith*,¹³⁹ is that control for minimum objective qualifications "ensure[s] that a plaintiff's methodology has eliminated the common nondiscriminatory explanation of a lack of qualifications."¹⁴⁰ While that rule may eliminate "the common nondiscriminatory explanation of a lack of qualifications," it does not eliminate the at least equally common nondiscriminatory explanation of a lack of *relative* qualifications.¹⁴¹

138. *Segar v. Smith*, 738 F.2d 1249, 1274 (D.C. Cir. 1984), *cert. denied sub nom. Meese v. Segar*, 471 U.S. 1115 (1985).

139. 738 F.2d 1249 (D.C. Cir. 1984), *cert. denied sub nom. Meese v. Segar*, 471 U.S. 1115 (1985).

140. *Id.* at 1274; *see also* *Palmer v. Shultz*, 815 F.2d 84, 91 n.6 (D.C. Cir. 1987) (quoting *Segar*).

141. The D.C. Circuit has dispensed with the need even to control for minimum qualifications in some cases. For circumstances where data concerning proxies are lacking, the court has erected "a rebuttable presumption of an equal distribution of qualifications between minority and majority group applicants." *De Medina v. Reinhardt*, 685 F.2d 997, 1008-09 n.7 (D.C. Cir. 1982); *see also* *Trout v. Lehman*, 702 F.2d 1094, 1102 n.10 (D.C. Cir. 1983) ("If data on a particular minimum objective qualification are simply unavailable, *Valentino's* requirement [that plaintiff control for 'minimum objective qualifications'] may be eased, for '[e]xactness is not required at the *prima facie* stage.'") (quoting *De Medina*, 686 F.2d at 1008-09 n.6), *vacated*, 465 U.S. 1056 (1984).

The Third Circuit, in the age discrimination case of *Bruno v. W.B. Saunders Co.*, 882 F.2d 760 (3d Cir. 1989), *cert. denied*, 493 U.S. 1062 (1990), also was willing to forgive an absence of control for minimum qualifications. The plaintiff introduced statistical studies demonstrating that employees over forty did not receive promotions at a rate proportionate to their representation in the work force. *Id.* The defendant argued that these studies were irrelevant because they did not take into account the minimum qualifications for the jobs. Relying on *Bazemore*, the court held that the studies were not inadmissible simply because they did not take

Under the *Segar* analysis, once the plaintiff controls for the minimum objective qualifications, the defendant bears the burden of providing evidence of other variables that should have been considered, unless the court concludes that the missing variable is so critical as to undermine the plaintiff's case completely. That mode of analysis would make sense only if employers hire or promote at random from a pool of applicants having the minimum qualifications for the job, but no one suggests that employers actually do, or even should, make decisions in that way.

The difficulty faced by even non-discriminating employers is exemplified by the analysis in *Mozee v. American Commercial Marine Serv. Co.*,¹⁴² a case involving, among other things, claims for racial discrimination in promotions to the position of "leadman." Plaintiffs had presented statistics demonstrating a modestly significant (2.13 standard deviations) disparity between promotions of blacks and whites from the hourly work force.¹⁴³ The defendant challenged the plaintiffs' analysis on the ground that it did not control for three variables that it contended were important to the promotion decision: (1) seniority; (2) membership in the craft of the open leadman position; and (3) first-class rank within that craft.¹⁴⁴ The defendant argued that once these qualifications were considered, the statistical disparities were reduced or eliminated. The trial court was thus faced with the question of whether the appropriate comparison for promotions to leadman was to the hourly work force as a whole or whether the comparison must be adjusted to take into account the above-described factors. The court rejected the defendant's attempt to limit the comparison, and the Seventh Circuit affirmed.¹⁴⁵

On appeal, the Seventh Circuit relied upon the Supreme Court's decision in *Bazemore v. Friday*¹⁴⁶ suggesting that plaintiffs need not

minimum qualifications into account. *Id.* at 766-67. The court thought it relevant that the case was not a pattern-or-practice case, but rather an individual disparate-treatment case in which the plaintiff was attempting to buttress her claim of discrimination with statistical evidence. *Id.* at 767. That distinction makes no difference, however, because in either event the statistical evidence is introduced to prove the same thing: that the employer has a general practice of discrimination. Evidence that does not demonstrate a pattern of discrimination in a pattern-or-practice case demonstrates no more merely because it is introduced in a different kind of case. As Judge Greenberg stated in dissent, "[t]he door is now open in this circuit to the admission of plainly irrelevant statistical evidence in individual disparate treatment cases." *Id.* at 776 (Greenberg, J., dissenting).

142. 940 F.2d 1036 (7th Cir. 1991), *cert. denied*, 113 S. Ct. 207 (1992).

143. *Id.* at 1042-43.

144. *Id.* at 1045.

145. *Id.* at 1046.

146. 478 U.S. 385 (1986).

take into account "all measurable variables" and that "failure to include variables will affect the analysis' probativeness, not its admissibility."¹⁴⁷ The court of appeals affirmed the district court's refusal to require the variables to be included, stating that "the use of hindsight to construct 'qualifications' for a position must be viewed with some suspicion" and that "[t]he trial judge was correctly dubious of the seductive logic of post-hoc explanation."¹⁴⁸ The court's statements reflect a failure to understand the nature of the employer's defense in a case like this. When an employer argues that a variable should be taken into consideration, the employer is not necessarily saying that the trait is essential to the job; instead, the employer is merely saying that it is relevant.¹⁴⁹ If the trait is relevant, promotion decisions are not independent of it, and, if it is not found with the same frequency in different groups, failure to control for the trait will incorrectly ascribe statistical differences to discrimination.¹⁵⁰

The *Mozee* court failed to understand the difference between qualifications—which are attributes that are relevant to the job—and minimum qualifications—which are absolute requirements.¹⁵¹ For

147. *Id.* at 400.

148. *Mozee*, 940 F.2d at 1045.

149. When an employer requires a minimum score of 50 on a written test and one applicant scores a 60 and the other scores a 70, the employer is likely to select the employee who scored a 70 because of that score; nonetheless, the employer's identification of the score of 70 might be deemed "post hoc." See Smith & Abram, *supra* note 13:

In selecting explanatory variables for inclusion in a regression analysis, the list of factors should not be limited to qualifications established by an employer as job requirements, and proved to be job related to the satisfaction of the legal standard of proof. A profit-maximizing employer, attempting to extract a return based on a person's estimated productivity, will consider not only those qualifications identified as job prerequisites, but all factors influencing an individual's productivity.

Id. at 66–67 (footnote omitted); see also James Gwartney et al., *Statistics, the Law and Title VII: An Economist's View*, 54 NOTRE DAME LAW. 633 (1979):

Regardless of whether or not the firm has a formal wage structure that rewards [certain] variables, if, for example, years of craft experience (or schooling) and earnings are consistently linked, independent of race, the statistics support the contention that the skills of employees with more craft experience or schooling are important determinants of job success for this firm.

Id. at 658.

150. See Smith & Abram, *supra* note 13, at 69–70: "When minorities (or females) have lower qualifications on the average than nonminorities (or males), regression analysis based on productivity variables will produce systematically biased and inflated race/sex coefficients suggesting the presence of discrimination when none exists."

151. Similarly, in *James v. Stockham Valves & Fittings Co.*, 559 F.2d 310, 332 (5th Cir. 1977), *cert. denied*, 434 U.S. 1034 (1978), the court disallowed use of a "years of schooling" variable in a salary regression because "education is not a job requirement" and because "white employees at Stockham have more education than blacks." The former reason is similar to the court's reasoning in *Mozee* in its confusion of minimum qualifications and productivity-related variables. As to the second reason, if white employees did not have more education than blacks,

example, the defendant argued that most foremen took into account a candidate's ability to read a blueprint in making leadman decisions. The district court, however, disbelieved this assertion because one of defendant's witnesses "testified that such a skill was not a leadman requirement."¹⁵² Similarly, the district court rejected the defendant's contention that the most important factor in the promotion decision was "experience as a first-class employee in the particular craft in which the vacancy occurred,"¹⁵³ a consideration that is highly plausible on its face in promoting a person to a position entailing supervisory responsibilities. The court rejected this factor because nearly ten percent of those selected for promotion did not have such experience.¹⁵⁴ Rather than diminishing the defendant's argument, however, this fact would seem to enhance it. After all, the defendant had presented strong evidence that this facially plausible factor was relevant to its decision by demonstrating that the qualification was present in over ninety percent of promotions.¹⁵⁵

Finally, the district court rejected the defendant's attempt to control for seniority. In affirming that rejection, the Seventh Circuit relied on the fact that seniority was not a "requirement" of the job and that the most senior employee did not always get the promotion.¹⁵⁶ Again, because the employer did not use strict seniority as the basis for promotions, the court deemed seniority irrelevant to the decision. However, the mere fact that an employer rejects a strict seniority approach to promotions does not mean that the amount of experience that an employee has is unrelated to his chances of success under a nondiscriminatory system of promotion; instead, it simply means that the

the issue would never have come up because inclusion of the variable of education would not have reduced the apparent race disparity. The only variables that are important in a regression to prove discrimination are precisely those that *are* correlated with race, sex, or other prohibited classifications.

152. *Mozee*, 940 F.2d at 1045.

153. *Id.* at 1046 (quoting the Appellant's Brief at 26).

154. *Id.*

155. The Seventh Circuit, in a puzzling footnote, seemed to suggest that if the selection system is a subjective one, the employer's ability to rely on objective criteria is diminished. *Id.* at 1048 n.11 ("The alleged prerequisites to promotion—certain skills, rank, and seniority—would not constitute common nondiscriminatory reasons for denying a promotion in a system employing purely subjective methods.").

156. The court stated:

Jeffboat also offered employee seniority as a *qualification* a foreman would consider in selecting a new or temporary leadman. . . . Like blueprint reading, this "*requirement*" was disavowed by defendant's own witness. More significantly, African-American employees were senior to the white employees actually promoted in eight of the forty-three promotions to permanent leadman.

Id. at 1046 (emphasis added; citations to record omitted).

most senior employee will not be promoted automatically.¹⁵⁷ Furthermore, failing to control for seniority will systematically prejudice many employers; because there is in many work forces a relationship between seniority and race,¹⁵⁸ seniority effects will be interpreted erroneously to be race effects.

Once a court recognizes the necessity of accounting for minimum qualifications, as most courts have done, it should follow that it is similarly necessary to account for all the major variations in productivity that may affect the likelihood of a person's being selected for a job. It seems inconsistent to insist, on the one hand, that general population statistics are inappropriate because skills are not randomly distributed throughout the population with respect to race, sex, and age, but to presume, on the other hand, that once minimum qualifications are accounted for the remaining productivity-related variables *are* randomly distributed. In fact, one would expect just the opposite — that the same sorts of systematic differences that we see in minimum qualifications exist with respect to other qualifications as well, whether or not they are measurable.¹⁵⁹ In short, so long as the courts fail to follow their reasoning to its logical conclusion, discrimination frequently will be found where none exists.

2. *Exclusion of Subjective or Otherwise Unquantifiable Factors*

Closely linked to the preference of courts for minimum qualifications in statistical analyses is their focus on "objective" qualifications. For example, in *Segar v. Smith*, the court held that the plaintiffs need not factor in the qualification of "specialized experience" for promotion because plaintiffs "were not realistically able to account for the

157. One would not expect there to be a linear relationship between seniority and promotion even if there is a substantial relationship between the two. An employee with six months experience is unlikely to be as qualified for promotion to leadman as an employee with six years experience. On the other hand, an employee with thirty years experience is likely to be less qualified for promotion than an employee with ten years experience, since if the employee has not been promoted to leadman in three decades, it is unlikely that he is leadman material.

158. Minorities as a group will commonly have less seniority than nonminorities as a group for a number of reasons, including: (1) the pool of qualified minority employees is increasing; (2) employers are engaging in more vigorous affirmative action efforts than in the past; (3) minorities tend to be younger than whites; and (4) the employer may have discriminated in the past but stopped. When it comes to providing layoff protection for minorities under a last-hired first-fired system, the fact that minorities will have lower than average seniority is not controversial. See *Wygant v. Jackson Bd. of Educ.*, 476 U.S. 267 (1986).

159. See *BALDUS & COLE*, *supra* note 39, § 9.02 ("[I]t cannot be assumed as a matter of fact that other unknown but potentially influential qualifications are uniformly distributed between the minority and majority group applicants."). But see *De Medina v. Reinhardt*, 686 F.2d 997, 1003 (D.C. Cir. 1982) (establishing a rebuttable presumption of an equal distribution of relevant skills).

application of so amorphous a criterion.”¹⁶⁰ According to the court, “[t]he appropriate degree of refinement of the plaintiffs’ statistical analysis . . . may depend on the quality and control of the available data.”¹⁶¹ Under the circumstances of that case, plaintiffs could not have quantified the qualification “in a manner that would have made the requirement amenable to statistical analysis.”¹⁶² Because the criterion was not quantifiable, if the defendant wanted to control for it in its own statistical rebuttal, it was obligated to distill “objective proxies” that were quantifiable.¹⁶³ According to the court, “[b]oth the policies underlying Title VII and general principles of evidence suggest that the burden of production of such evidence must rest with the defendant.”¹⁶⁴

The *Segar* court’s imposition of the burden of production on the defendant was inappropriate in a regime that places the persuasion burden on the plaintiff. By definition, imposing a burden of persuasion on a party means that that party bears the risk if evidence on a point is lacking. In *Segar*, the court held that lack of evidence on the point (i.e., the inability to quantify the qualification) was sufficient justification for shifting the burden of production to the employer.¹⁶⁵ In reality, however, shifting the burden of production to a party because evidence is lacking constitutes a shift not only in the burden of production, but in that of persuasion as well.¹⁶⁶ Because the *Segar* court had already determined that the qualification could not be quantified, shifting the burden of production to the defendant to quantify the unquantifiable was equivalent to shifting the burden of persuasion to the defendant to disprove equality of qualifications, a burden that, by definition, could not be carried.

160. *Segar v. Smith*, 738 F.2d 1249, 1276 (D.C. Cir. 1984), *cert. denied sub nom. Meese v. Segar*, 471 U.S. 1115 (1985).

161. *Id.* (quoting *Trout v. Lehman*, 702 F.2d 1094, 1101 (D.C. Cir. 1983)).

162. *Id.*

163. *Id.*

164. *Id.*

165. *See also De Medina v. Reinhardt*, 686 F.2d 997, 1008–09 n.7 (D.C. Cir. 1982) (imposing burden on employer to disprove equality of qualifications “when data are unavailable”).

166. The burden of persuasion is important in two somewhat different circumstances. The first is where there is evidence on both sides of an issue and the trier of fact judges the evidence to be in equipoise; only when the weight of the evidence is judged to be exactly 50/50 does the civil burden of persuasion decide the case, because regardless of who bears the burden of persuasion a 51/49 split either way will result in a favorable decision for the party with the 51 percent. The second, and more important, class of cases is where evidence on a point is lacking; in these cases, the party with the burden of persuasion automatically loses. It is in these cases that shifting the burden of production because evidence is lacking is equivalent to shifting the burden of persuasion.

The focus on quantifiable traits has created a Procrustean bed that distorts both the way cases are litigated and, perhaps even more importantly, the way employment decisions are made. Courts originally accepted statistical analysis in discrimination cases because of a belief that the kind of evidence available in such cases lent itself to statistical analysis.¹⁶⁷ Statistical analysis, having become accepted, is now driving the kind of evidence deemed relevant or necessary. However, if evidence is not amenable to statistical analysis—that is, if statistical analysis does not further the search for truth in the context of a particular set of facts—the method of analysis should be altered, rather than one's judgment about what is true. The fact that statistical evidence is the only evidence that exists is not a basis for saying that it is enough.¹⁶⁸

Insistence upon quantification puts employers who rely on non-quantitative factors at significant risk. If they rely upon non-quantitative factors in making decisions (a practice particularly appropriate in salary and promotion cases, where the employer has had significant opportunity to evaluate the employee's performance, usually by a process that includes a significant subjective component), they run the risk of not being able mathematically to justify any resultant disparities.¹⁶⁹ This imposes significant pressure on employers either to increase reliance on quantitative factors in making employment decisions—even if those factors are less likely to yield the optimal decision—or to avoid statistical disparities altogether by engaging in quota hiring.

In one of the early articles on mathematical proof in litigation, Laurence Tribe eloquently warned of one of the significant pitfalls of using statistical proof—what he calls “the dwarfing of soft variables”:

167. See *Coates v. Johnson & Johnson*, 756 F.2d 524, 540 (7th Cir. 1985) (“[C]ourts have sometimes relied on statistical evidence as the best means of showing the cumulative effects of employment actions.”); *CONNOLLY ET AL.*, *supra* note 9, § 1.02 (“Statistical analysis is well suited to discrimination cases, since it essentially involves making comparisons among classes of people.”).

168. *But see Capaci v. Katz & Besthoff, Inc.*, 711 F.2d 647, 653 (5th Cir. 1983) (reversing district court and rejecting the defendant's argument that the plaintiffs' statistical comparison was too crude, stating that “[t]he defendant would require refinements beyond that available in published statistics”), *cert. denied*, 466 U.S. 927 (1984); *Trout v. Lehman*, 702 F.2d 1094, 1101 (D.C. Cir. 1983) (“The appropriate degree of refinement of the plaintiffs' statistical analysis, moreover, may depend on the quality and control of the available data”), *vacated*, 465 U.S. 1056 (1984); Cohen, *supra* note 52, at 387 (pointing out that because employers engage in discrimination surreptitiously, “[t]he most common (and often the only) evidence in these cases consists of empirical data about the available labor pool and the work force”).

169. For a defense of the use of subjective factors in making employment decisions, see Browne, *supra* note 79, at 334–38.

The syndrome is a familiar one: If you can't count it, it doesn't exist. Equipped with a mathematically powerful intellectual machine, even the most sophisticated user is subject to an overwhelming temptation to feed his pet the food it can most comfortably digest. Readily quantifiable factors are easier to process—and hence more likely to be recognized and then reflected in the outcome—than are factors that resist ready quantification. The result, despite what turns out to be a spurious appearance of accuracy and completeness, is likely to be significantly warped and hence highly suspect.¹⁷⁰

The syndrome described by Tribe is a pervasive one in discrimination cases, and it is more distorting in discrimination cases than it is in criminal cases, which were the primary focus of his article. In criminal cases, the concern is that the trier of fact, although having all of the relevant evidence before it, will tend to place undue weight on the apparently scientific quantitative evidence and discount or disregard the non-quantitative evidence. In discrimination cases, however, the concern is not just that the trier of fact will place undue weight on the quantitative evidence, but that the non-quantitative evidence will not even be admitted. In *Segar*, the court did not simply hold that the quantitative analysis was sufficiently strong as to dwarf the non-quantifiable qualification; it held that absent quantification the qualification had no explanatory value and was therefore appropriately excluded by the trial court.

Focus on the easily quantifiable aspects of the labor force obscures what are often more important variables. For example, no one seriously believes that all persons with a given length of education are equally educated—that high school graduates, or college graduates, or law school graduates are all fungible. One year of education is not necessarily equivalent to another year of education. Schools differ in what they teach and how well they teach it, and students differ in terms of how well they learn and in what they learn.¹⁷¹ These critical variations in educational experience, however, are declared irrelevant to the plaintiff's *prima facie* case.

170. Tribe, *supra* note 4, at 1361–62. Brilmayer and Kornhauser, *supra* note 39, at 149, suggest that Tribe's arguments against statistical methods "may go too far," since they might eliminate the use of statistics in Title VII cases. However, the proper measure of whether an argument goes too far is not how many cases the argument affects, but how many cases the argument *inappropriately* affects. If statistics in Title VII cases suffer from the same flaws as statistics in other kinds of cases, one should properly hesitate to use them.

171. See generally Jere R. Behrman & Nancy Birdsall, *The Quality of Schooling: Quantity Alone is Misleading*, 73 AM. ECON. REV. 928 (1983). Justice Stevens relied upon this obvious fact in his concurring opinion in *Washington v. Davis*, 426 U.S. 229, 254 (1976), in which he observed that an employer's use of standardized tests was justified in light of the fact that there is a substantial disparity between the quality of various schools.

The focus on quantification creates pressure to quantify variables for quantification's sake, a pressure resisted by the majority in *EEOC v. Sears, Roebuck & Co.*¹⁷² but not by the dissent. The EEOC had accused Sears of sex discrimination in hiring commission salespersons, and Sears defended in part on the ground that men and women were not equally interested in such positions. Sears presented extensive, uncontradicted testimony of numerous store managers and personnel managers concerning their unsuccessful efforts to recruit women into these positions, and it introduced job-interest surveys and polls that supported its position that women were substantially less interested in commission sales than men. The EEOC attempted to deal with this testimony by incorporating into its statistical analysis an interest-factor adjustment—under which the proportion of men interested in commission sales was assumed to be three times as great as the proportion of women so interested. Because the EEOC had not produced any evidence to establish that a factor of three was of the appropriate magnitude, the court held that the district court was correct to reject that analysis.¹⁷³

In dissent, however, Judge Cudahy followed an approach similar to that of the court in *Segar*, suggesting that the EEOC's interest-factor adjustment should have been accepted because of the "skepticism that courts ought to show toward defenses to Title VII actions that rely on unquantifiable traits ascribed to protected groups."¹⁷⁴ He asserted that "a critique based on such obviously unquantifiable and peripheral considerations is inordinately critical of the statistical evidence being presented" and that because the EEOC had not failed to include any *measurable* variables, it should not lose simply because it had not "disprov[ed] the enormous significance that Sears attributes to *unmeasurable* variables."¹⁷⁵ Thus, despite the fact that the EEOC had simply pulled the number "3" out of a hat, Judge Cudahy would have imposed on the defendant the obligation of proving that the number was too small. Given his recognition that the variable was not measurable, however, the placement of the burden of proof would be dispositive and the defendant's failure to carry its burden would be foreordained.

172. 839 F.2d 302 (7th Cir. 1988).

173. *Id.* at 335.

174. *Id.* at 361 (Cudahy, J., concurring in part and dissenting in part).

175. *Id.* at 363. The Ninth Circuit has announced that it "agree[s] with the dissent and reject[s] the approach taken by the *Sears* majority which places a very heavy—and possibly insurmountable—burden on the plaintiff with respect to establishing the probativeness of proffered statistical data." *EEOC v. General Tel. Co. of Northwest*, 885 F.2d 575, 581 (9th Cir. 1989), *cert. denied*, 111 S. Ct. 370 (1990).

Courts should abandon their suspicion of non-quantifiable variables. Where such variables are relevant, as they often are, courts should recognize that the non-quantitative nature of the relevant evidence makes quantitative analysis inappropriate, rather than concluding that the plaintiffs' decision to rely on quantitative analysis invalidates the defendant's non-quantitative evidence.¹⁷⁶

3. *Omission of Variables*

In many, if not most, cases where plaintiffs attempt to make a statistical showing of discrimination, the defendant argues that the plaintiff has not compared the appropriate statistics or has failed to take into account some important variable—such as a particular qualification or differential interest—that the defendant maintains would (or might) make a difference. The important question, as yet not fully resolved, is what the consequence of that failure is and who bears the burden in those circumstances. The leading case is *Bazemore v. Friday*.¹⁷⁷ In *Bazemore*, the lower court had held the plaintiffs' statistical evidence inadmissible because of the plaintiffs' failure to control for a particular variable asserted by the defendant to be an important one.¹⁷⁸ The Supreme Court rejected this conclusion, stating:

While the omission of variables from a regression analysis may render the analysis less probative than it otherwise might be, it can hardly be said, absent some other infirmity, that an analysis which accounts for the major factors "must be considered unacceptable as evidence of discrimination." Normally, failure to include variables will affect the analysis' probativeness, not its admissibility.¹⁷⁹

Relying on *Bazemore*, numerous courts have held that omission of a variable from a regression does not undermine the plaintiff's prima facie case. Rather, if the defendant wants to impugn the plaintiff's statistical showing, it must introduce evidence to show that the miss-

176. See Brilmayer & Kornhauser, *supra* note 39, at 153 ("[S]ince legal problems are subtle and complex, the unquantifiable variables may well dwarf the quantifiable ones and make numerical modelling futile."); Campbell, *supra* note 41, at 1303:

It is entirely acceptable within economics, and other social sciences, to use a model to help explain observed behavior where no other model works better, even though in an absolute sense, the model does not describe the data very well at all. But this cannot be the attitude . . . for proof of legal propositions.

177. 478 U.S. 385 (1986).

178. In their salary regression, plaintiffs had accounted for race, education, tenure, and job title, but they had not controlled for county of employment. *Id.* at 398.

179. *Id.* at 400 (quoting *Bazemore v. Friday*, 751 F.2d 662, 672 (4th Cir. 1984)). The Court then went on to acknowledge that some statistical studies are "so incomplete as to be inadmissible as irrelevant." *Id.* at 400 n.10.

ing variable would have made a difference.¹⁸⁰ Courts employing such reasoning, however, are misreading *Bazemore*.

The standard reading of *Bazemore* seems to be that in order for the defendant to dispel the prima facie case by pointing to the plaintiff's failure to control for a relevant variable, the defendant must show that the plaintiff's statistical model is so flawed as to be inadmissible.¹⁸¹ Otherwise, it is the defendant's obligation to account for the missing variable and, failing that, the defendant's attempt to rebut the prima facie case is deemed unsuccessful.¹⁸² However, that is not what *Bazemore* held. *Bazemore* held only that as long as the "major fac-

180. *Berger v. Iron Workers Reinforced Rodmen Local 201*, 843 F.2d 1395, 1416 (D.C. Cir. 1988) ("[T]o demonstrate that statistical evidence is invalid, the challenger must present 'credible evidence that the statistical proof is defective' and 'a plausible explanation of how the asserted flaw is likely to bias the results against his or her position.'"), *cert. denied*, 490 U.S. 1105 (1989); *Palmer v. Shultz*, 815 F.2d 84, 101 (D.C. Cir. 1987) ("The logic of *Bazemore* . . . dictates that in most cases a defendant cannot rebut statistical evidence by mere conjectures or assertions, without introducing evidence to support the contention that the missing factor can explain the disparities as a product of a legitimate, nondiscriminatory selection criterion."); *Trout v. Lehman*, 702 F.2d, 1094, 1102 (D.C. Cir. 1983) ("[U]nquantified, speculative, and theoretical objections to the proffered statistics are properly given little weight by the trial court."), *vacated*, 465 U.S. 1056 (1984).

181. *See, e.g., Berger*, 843 F.2d at 1413 (suggesting that defendants can rebut the prima facie case by articulating a legitimate nondiscriminatory explanation for the disparity, by introducing alternative statistical evidence, or "by demonstrating that plaintiffs' statistics are so flawed as to be meaningless").

In *Penk v. Oregon State Board of Higher Education*, 816 F.2d 458 (9th Cir.), *cert. denied*, 484 U.S. 853 (1987), the Ninth Circuit gave some indication of what it views as sufficient flaws as to deprive statistics of substantial probative value. In *Penk*, the court held that the district court did not clearly err when it discounted the plaintiffs' statistical analysis in an action challenging the salary, promotion, and tenure practices in the Oregon higher education system because the analysis omitted variables such as teaching quality, community and institutional service, and quality of research and scholarship. *Id.* at 465. Similarly, in *Sheehan v. Purolator, Inc.*, 839 F.2d 99, 106 (2d Cir.), *cert. denied*, 488 U.S. 891 (1988), the court held that the district court had not clearly erred in finding plaintiffs' statistical analysis, which had been admitted into evidence, "flawed" for failing to take into account education, prior work experience, and job level. It is difficult to understand on what theory such obviously meaningless statistical studies were deemed sufficiently relevant even to be admitted into evidence.

182. *See, e.g., EEOC v. General Tel. Co. of Northwest*, 885 F.2d 575, 579 (9th Cir. 1989) ("*Bazemore* requires that the defendant do more than simply point out possible flaws in the proponent's statistical analyses in order to rebut the inference of discrimination raised by the statistical evidence."), *cert. denied*, 111 S. Ct. 370 (1990); *Sobel v. Yeshiva Univ.*, 839 F.2d 18, 33-34 (2d Cir. 1988), *cert. denied*, 490 U.S. 1105 (1989) (Even though the district court had ruled that flaws in plaintiffs' regression analysis affected its weight rather than its admissibility, the court of appeals chided the district court for failing to understand that *Bazemore* "require[s] a defendant challenging the validity of a multiple regression analysis to make a showing that the factors it contends ought to have been included would weaken the showing of a salary disparity made by the analysis."); *Palmer v. Shultz*, 815 F.2d 84, 101 (D.C. Cir. 1987) ("[I]n most cases a defendant cannot rebut statistical evidence . . . without introducing evidence to support the contention that the missing factor can explain the disparities as a product of a legitimate, nondiscriminatory selection criterion.").

tors” were considered by the analysis, failure to control for other factors did not render the analysis *inadmissible*. But to say that evidence is admissible is not the same as saying that it is strong or, *a fortiori*, that it is sufficient to establish a *prima facie* case. For example, evidence that a person was in the United States when a murder was committed in Arkansas is relevant and probative and is surely admissible. No one would argue, however, that it is strongly suggestive that the person committed the crime. Rather, it is simply background evidence that is consistent—or at least not inconsistent—with guilt. *Bazemore* provides no authority for the proposition that any evidence consistent with discrimination establishes a *prima facie* case, and there is no basis in *Bazemore* or in any other Supreme Court decision for the proposition that any probative statistical evidence, no matter how weak, suffices to establish a *prima facie* case.

Ideally, before courts draw a conclusion based upon a disparity between an employer’s work force and some comparative group, every possible variable that might be relevant to selection would be controlled for.¹⁸³ In practice, however, evidence is never perfect, and all conceivable variables cannot be controlled for. The important practical question is how far may the evidence deviate from ideal and still be deemed to retain substantial probative value. Unfortunately, the answer for too many courts, based upon their misreading of *Bazemore*, is “quite a lot.”

Courts have been willing to endorse statistical showings based upon assumptions that are wholly counterintuitive, rejecting defendants’ contrary assertions as mere “speculation.” For example, the Eighth Circuit held in *Catlett v. Missouri Highway and Transportation Commission*¹⁸⁴ that the fact only 10 percent of hires for highway maintenance workers were female, while 48 percent of the local work force was female, was suggestive of discrimination, despite the fact that the plaintiffs’ statistics failed to take into account the actual interest of qualified potential employees. The court employed the familiar burden-shifting approach and argued that Missouri bore the burden of introducing evidence to show that failure to account for interest was significant, because, according to the court, “[m]ere conjecture or assertion on [a] defendant’s part . . . cannot defeat the inference of discrimination created by [a] plaintiff’s statistics.”¹⁸⁵ Then, in response to the defendant’s argument that it hired female applicants at

183. See Fisher, *supra* note 41, at 709 (“Obviously, the assumption that one has controlled for all the important influences is basic to any attempt to measure those influences correctly.”).

184. 828 F.2d 1260 (8th Cir. 1987), *cert. denied*, 485 U.S. 1021 (1988).

185. *Id.* at 1266 (quoting *Palmer*, 815 F.2d at 101).

a slightly higher rate than it hired male applicants, the court had the temerity to state that this fact was irrelevant because "[v]ictims of a discriminatory policy cannot be told they have not been wronged because other females have been hired."¹⁸⁶

In rejecting defendants' challenges to plaintiffs' assumptions as "speculation" and "conjecture," courts have paid scant attention to the fact that the assumptions that defendants are challenging are no less speculative or conjectural.¹⁸⁷ In fact, the assumption of equal qualifications and interest is often contrary to common experience, as is neatly illustrated by *Catlett*, in which the court assumed that men and women were equally interested in highway maintenance jobs. The statistical comparisons contained in a plaintiff's analysis necessarily rest on a fundamental and critical assumption of equal interest and equal qualifications. If that assumption is inaccurate, the statistical analysis is meaningless.¹⁸⁸ Similarly, when a plaintiff uses a proxy for a productivity variable, the statistical model assumes that the proxy is an accurate representation of the missing variable itself; yet, courts often impose on the defendant the burden of proving that a plaintiff's chosen proxy is unrepresentative.¹⁸⁹ Given that plaintiffs bear the burden of proof on the ultimate question of discrimination, it is hard to understand why the defendant should bear the burden of proof of the falsity of the plaintiffs' assumptions.

Several potential reasons exist for the placement of the burden on the defendant to prove the relevance of missing variables. First, the assumption of equal interest and qualifications could be the efficient starting assumption. However, the assumption of equal interest and

186. *Id.*

187. See *Sobel v. Yeshiva Univ.*, 839 F.2d 18, 35 (2d Cir. 1988) (Defendant "cannot rely on assumptions about imperfections inherent in productivity proxies, nor can it simply propose alternative variables without justifying their inclusion."), *cert. denied*, 490 U.S. 1105 (1989).

188. Baldus and Cole assert that "[i]f the law assumes equal eligibility . . . there is no problem [with the validity of the statistical analysis]. But when it does not, there are distinct threats to validity." BALDUS & COLE, *supra* note 39, § 9.02. They define "validity" as "the property of measuring what the figures purport to measure and not some other factor." *Id.* § 2.312. However, to say that a counterfactual assumption preserves the "validity" of a statistical analysis is a strange use of the term. It may be that the assumption of equal qualifications, whether or not it is well founded, makes a statistical comparison legally relevant; that is, one may say that because equal qualifications are presumed, a significant difference between two groups establishes a prima facie case. It does not mean, however, that the statistical analysis "measur[es] what the figures purport to measure." The statistical comparison purports to measure, depending upon the particular court's formulation, either the likelihood that the disparity was caused by chance or the likelihood that it was caused by discrimination; the accuracy of the answer is not enhanced by changing the underlying factual assumptions. In other words, the law cannot assume away threats to validity; it can only make threats to validity legally irrelevant.

189. See *Sobel*, 839 F.2d at 34-35.

qualifications is so often wrong that efficiency is not a reason to retain it. Race, sex, and age differences in interest, qualifications, and ability are common enough that the party bearing the burden of persuasion should also bear the burden of production. Even if a production burden were appropriate, the defendant should bear only the obligation of presenting evidence creating a genuine issue of fact that the assumption of equality is unwarranted, a burden much lighter than that seemingly imposed by most courts.¹⁹⁰ Even the *Sears* court, for example, seemed to view the question before it as whether Sears' interest evidence had eliminated the apparent disparities—that is, whether Sears had satisfied a persuasion burden¹⁹¹—when all it should have required is that Sears demonstrate that a genuine issue of fact existed concerning the correctness of the equal-interest assumption, which Sears could have done with substantially less evidence than it provided.

The second reason for placing the burden on the defendant to demonstrate the falsity of the plaintiff's assumptions is that defendants have superior access to evidence, a rationale which, at most, would justify a production burden rather than a persuasion burden. The assumption of superior access, however, is unfounded for several reasons. First, in many cases, evidence concerning interests and qualifications of various populations must be generated after the litigation begins from data not in the possession, or at least not in the exclusive possession, of the defendant.¹⁹² Plaintiffs can generate it as well as defendants, and the fact that defendants may enjoy superior financial resources is not a basis for shifting evidentiary burdens. Second, after discovery in the typical class action, the plaintiffs have as much evidence concerning the profile of employees and applicants as the defendant.¹⁹³ Third, the fact that defendants as a class may have supe-

190. In *Palmer v. Shultz*, 815 F.2d 84, 106–07 n.19 (D.C. Cir. 1987), the court rejected the defendant's argument that the plaintiffs' statistical case was flawed by its failure to control for differential interest in the jobs at issue where the defendant "presented no evidence *at all* that *preference* would *explain* the disparities related to sex." The court did not discuss what the consequence of the defendant's introduction of some evidence of differential interest would be.

191. *Sobel*, 839 F.2d at 334–35 ("[W]e conclude that the district court did not clearly err in determining that Sears' interest evidence substantially reduced (and indeed almost eliminated) the EEOC's alleged promotion disparities.").

192. For example, in *De Medina v. Reinhardt*, 686 F.2d 997, 1009 n.7 (D.C. Cir. 1982), the court suggested that possible sources of evidence of unequal qualifications were "data on the qualifications of applicants processed by similarly situated decision makers" and "published work force or census data." See also *BALDUS & COLE*, *supra* note 39, § 6.2, at 195.

193. Richard Lempert has argued that in *Melani v. Board of Higher Education*, 561 F. Supp. 769 (S.D.N.Y. 1983), a case alleging sex discrimination in salaries at City University of New York, the defendant should have borne the burden of introducing evidence that the salary discrepancies were due to the fact that women were found disproportionately in departments that had low salary scales (such as education), while men were found disproportionately in

rior access to a kind of evidence may not warrant a presumption that each individual defendant has such access. If, for example, a particular form of evidence exists in only 20 percent of cases, but when it does exist it is usually found in the defendants' employment records, one could say that in some sense defendants as a class have "superior access" to it. However, unless the circumstances of a particular case suggest that the defendant in fact has the evidence, there is no reason to count the absence of evidence against the defendant in a regime that places the risk of nonpersuasion on the plaintiff.

The third reason for requiring the defendant to disprove the plaintiff's assumption of equal interest and qualifications could be that such a rule is more consistent with the substantive equality goals of Title VII than an assumption of unequal interest and qualifications.¹⁹⁴ However, Title VII did not legislate away differences between the races and the sexes, nor could it; rather, it prohibited discrimination based upon race and sex. Furthermore, Title VII did not mandate that employers hire without regard to qualifications and interest; it did just the opposite.¹⁹⁵ The central tenet of Title VII is not that there are no differences between men and women or between blacks and whites; it is that maleness, femaleness, blackness, and whiteness ordinarily are not to be the basis for employment decisions. Given that premise, there is no basis for suggesting that Title VII requires that courts presume that which they know not to be true.¹⁹⁶

departments having higher scales (such as engineering). Lempert, *supra* note 39, at 1108-09. Relying on the notion that "failure to respond to an opponent's argument is itself an important piece of information," he suggests that if the defendant, having raised the possibility that academic department explains the disparity, "fails to test for it, the defendant's failure is itself reason to believe that a consideration of this plausible nondiscriminatory explanation would not exonerate the defendant." *Id.*

Lempert implies that the information costs of the parties would be markedly different; yet in order for the plaintiffs to have made their initial *prima facie* showing, they must have already had salary information by sex and probably also by department. Whether or not they had department-by-department information, they surely could have obtained it through discovery. Lempert does not explain why, if plaintiffs had data concerning a variable that admittedly is a "plausible nondiscriminatory explanation" for the disparity, the plaintiffs' failure to introduce that evidence should not be held against *them*.

194. See *De Medina*, 686 F.2d at 1003-09 n.7 ("Both equitable considerations and, in Title VII cases, the policy of the statute, support a rebuttable presumption of an equal distribution of qualifications between minority and majority group applicants when data are unavailable.").

195. As the Supreme Court stated in *Griggs v. Duke Power Co.*, 401 U.S. 424, 436 (1971): "Congress has not commanded that the less qualified be preferred over the better qualified simply because of minority origins. Far from disparaging job qualifications as such, Congress has made such qualifications the controlling factor, so that race, religion, nationality, and sex become irrelevant."

196. The court in *De Medina*, 686 F.2d at 1009 n.7, gave an additional reason for imposing the burden on the employer: "[I]t was the defendant's selection process that produced the

4. "Tainted Variables"

Treatment of factors alleged to have been tainted by discrimination is another way in which some courts have improperly shifted the burden of proof to defendants in pattern-or-practice cases. This issue arises primarily in salary and promotion cases—cases in which the employer has had some prior opportunity to engage in discrimination against the affected employees. Typically either the employer argues that the plaintiff's regression analysis should be rejected or minimized because it omits an important variable and the plaintiff in turn asserts that the omitted variable is tainted by the employer's discrimination, or the plaintiff argues that a defendant's regression should be rejected because it has included such a variable.

There are two polar ways to deal with the "tainted variable" issue. First, omission of the variable could be viewed as weakening the plaintiff's statistical analysis unless the plaintiff can prove that the omitted variable is in fact tainted by discrimination. Second, the plaintiff's assertion of discrimination in the omitted variable could shift the burden to the defendant to disprove discrimination in the variable. Courts have differed in their approach to the "tainted variable" issue but virtually all, to some extent, have been willing improperly to relieve the plaintiff of the burden of proving something that should be the plaintiff's obligation to prove.

Some courts have expressly placed on the defendant the burden of disproving discrimination in the challenged variable. For example, in *Valentino v. United States Postal Service*,¹⁹⁷ the D.C. Circuit announced that "[a]bsent clear, affirmative evidence that promotions were made in accordance with neutral, objective standards consistently applied, there is no assurance that level or rank is an appropriate explanatory variable, untainted by discrimination."¹⁹⁸ As a result, the court stated that it did not consider the plaintiff's salary regression flawed because of its failure to include grade level as an explanatory variable.¹⁹⁹ Similarly, in *James v. Stockham Valves & Fittings Co.*,²⁰⁰ a case involving allegations of wage discrimination, the court considered a defense regression analysis that took into account "skill level" and

observed disproportionate impact, and it is he who is alleging that it is the product of differential qualifications among the applicants." The court seemed to lose sight of the fact that the issue is whether the statistical evidence is sufficiently powerful as to create a prima facie case; it is always the plaintiff's obligation to put on evidence with that power.

197. 674 F.2d 56 (D.C. Cir. 1982).

198. *Id.* at 73 n.30.

199. *Id.* at 71 n.26.

200. 559 F.2d 310 (5th Cir. 1977), *cert. denied*, 434 U.S. 1034 (1978).

"merit rating."²⁰¹ The court rejected the regression because the skill level could have been tainted by discrimination in job assignments and the merit rating could have been tainted by discrimination in performance evaluations.²⁰² Not surprisingly, once the defendant was not permitted to defend wage disparities on the basis of skill and merit, it could not rebut the prima facie case.

Other courts have taken what at first glance appears an opposite approach, requiring the plaintiff to prove discrimination in the omitted variable as a condition of excusing its omission. For example, in *Presseisen v. Swarthmore College*,²⁰³ in which plaintiffs raised claims of sex discrimination in hiring, promotion, and salary, the court rejected plaintiffs' multiple regression study introduced to prove salary discrimination on the ground that the study failed to take into account academic rank.²⁰⁴ Plaintiffs had asserted that academic rank had been excluded because the college had discriminated in promotions.²⁰⁵ The court concluded that plaintiffs had failed to demonstrate discrimination in promotions, however, because the difference in time to promotion for men and women was not statistically significant.²⁰⁶ As a result, the court concluded that academic rank was not tainted by discrimination and therefore should have been included, although the opinion implies that if the difference in time to promotion between men and women had been statistically significant it would have inferred discrimination from that fact alone and excluded the variable.²⁰⁷

The Seventh Circuit has taken a compromise position on the question of burdens. In *Coates v. Johnson & Johnson*,²⁰⁸ the court held that a plaintiff need not account for a "potentially biased factor" in estab-

201. The "skill level" variable was based on the employee's job class and was primarily a measure of quality of work experience with the defendant. *Id.* at 332. The "merit rating" variable was based upon subjective evaluations of employees by their supervisors. *Id.*

202. *Id.* ("If there is racial bias in the subjective evaluations of white supervisors, then that bias will be injected into [defendant's] earnings analysis.").

203. 442 F. Supp. 593 (E.D. Pa. 1977), *aff'd mem.*, 582 F.2d 1275 (3d Cir. 1978).

204. *Id.* at 614.

205. *Id.*

206. *Id.* at 612.

207. *Id.* at 614; see also *Sobel v. Yeshiva Univ.*, 839 F.2d 18, 35 (2d Cir. 1988) (upholding inclusion of academic rank in defense regression in light of the district court's finding that there was no sex discrimination in promotions), *cert. denied*, 490 U.S. 1105 (1989); *Ottaviani v. State Univ.*, 875 F.2d 365, 375 (2d Cir. 1989) (upholding district court's inclusion of rank variable where plaintiffs introduced no statistical evidence of discrimination in rank and defendants introduced "persuasive objective evidence to demonstrate that there was no discrimination in either placement into initial rank or promotion"), *cert. denied*, 493 U.S. 1021 (1990).

208. 756 F.2d 524 (7th Cir. 1985).

lishing its prima facie case,²⁰⁹ but that if the defendant in rebuttal offers a statistical analysis using an allegedly biased factor, the plaintiff bears the burden of persuading the trier of fact that the factor is actually biased. The court recognized that to impose the ultimate burden on the employer to prove nondiscrimination in the questioned variable “would skew the general Title VII framework for allocating burdens and would be inconsistent with the principle that the plaintiffs in a Title VII case retain the ultimate burden of persuasion on the issue of discrimination.”²¹⁰ According to the court, “[p]laintiffs should not be able to shift the burden of persuasion by alleging that some factor in defendant’s control has been used by the defendant discriminatorily.”²¹¹

The courts in all of these cases were willing to allocate proof burdens improperly. In *Segar* and *Stockham Valves*, the courts showed this willingness in a rather obvious way by requiring the defendants to disprove discrimination as a condition precedent to incorporating obviously relevant variables into their analyses. In *Presseisen*, the flaw is less obvious because the court required the plaintiff to demonstrate discrimination in promotion before it would approve exclusion of academic rank from the regression. To that extent, then, *Presseisen* is more consistent with the general rules of Title VII proof burdens than the prior two cases, although the court’s apparent willingness to infer discrimination from statistically significant differences in time to promotion was unjustified. However, even if there were demonstrable discrimination in promotion in *Presseisen*, it does not follow that academic rank should be ignored in the salary regression. Discrimination in salary is an action distinct from discrimination in promotion; an employer might engage in one but not the other. The two separate questions are: (1) did Swarthmore College discriminate in promoting women to higher ranks?; and (2) did it discriminate against women in salary once they had achieved the higher rank? If Swarthmore had discriminated in promotions but not in salary, omission of the academic rank variable would result in the erroneous conclusion that there had been discrimination in salary.²¹² On the other hand, if

209. *Id.* at 544. Apparently, the court meant by “potentially biased factor” a factor for which there is some evidence of discrimination, even if there has been no finding that discrimination exists.

210. *Id.*

211. *Id.*

212. Of course, to the extent that the plaintiff is arguing that she was discriminated against in promotion and that one of the consequences of not receiving the promotion was that she did not obtain the higher salary that would have gone with the promotion, the salary increment is recoverable as a remedy for *promotion* discrimination.

Swarthmore *had* discriminated against women in salary, that fact would show up even if the rank variable were included. That is, the statistical evidence would show that it took women longer to become associate or full professors—i.e., that they were discriminated against in promotion—and that once they achieved those ranks they were paid less than men—i.e., that they were discriminated against in salary. Thus, even though the *Presseisen* court appeared sensitive to ensuring that the proper parties bore the burden of proof on the various issues, its implicit willingness to reject the rank variable if discrimination were found in promotions would have improperly allowed the plaintiffs to prevail on claims of discrimination in both promotions and salary when they had presented evidence supporting only the former.

Michael Finkelstein argues in just the opposite direction, suggesting that the court in *Presseisen* let the defendant off too easily.²¹³ He asserts that the court was wrong to conclude that there was no discrimination in promotions merely because there were no statistically significant differences in time to promotion.²¹⁴ Finkelstein is correct that failing to reject the null hypothesis of no discrimination is not the same as proving it; hypothesis testing can never prove the null hypothesis.²¹⁵ In fact, Finkelstein's point would be as correct if there were *no* differences in time to promotion between men and women; in the absence of discrimination perhaps women would be promoted faster than men rather than at the same rate. However, the question should not be whether we can say that there is zero chance that the employer has discriminated in the challenged variable. Even the courts in *Segar* and *Stockham Valves* probably would have accepted a showing by the defendant that there was no statistically significant difference between the groups at issue in the challenged variable.²¹⁶ Nonetheless, Finkelstein argues that rank should be included in such regressions "only when there is clear evidence of neutral and objective standards that have consistently been followed in granting rank, so that there is *no chance* for discrimination,"²¹⁷ standards he acknowledges that most

213. Michael O. Finkelstein, *The Judicial Reception of Multiple Regression Studies in Race and Sex Discrimination Cases*, 80 COLUM. L. REV. 737, 742 (1980).

214. See *id.* ("[A]n absence of statistically significant differences is not equivalent to affirmative evidence that promotions were made neutrally . . .").

215. See Campbell, *supra* note 41, at 1304; Kaye, *Numbers Game*, *supra* note 39, at 839. But see McKeown, *supra* note 136, at 659 (suggesting that a statistical study can disprove discrimination).

216. Of course, if there were not such a difference, the variable probably would not explain the observed disparity. This demonstrates the problem with inferring discrimination from statistical disparities; it is most likely to cause exclusion of variables in precisely those cases where the variables have the most explanatory power.

217. Finkelstein, *supra* note 211, at 742.

academic institutions have not developed. However, he does not explain how a rule that imposes on defendants an obligation to disprove discrimination under a standard apparently comparable to the criminal “beyond a reasonable doubt” standard is derived from a law that requires plaintiffs to bear the burden of persuasion and requires proof only by a preponderance of the evidence. In fact, his argument simply demonstrates how doctrines of statistical proof have been corrupted to justify gross, but covert, deviations from governing substantive and procedural rules.

The *Coates* court’s apparent sensitivity to proof burdens did not prevent it from permitting a plaintiff to omit an admittedly relevant variable as part of its *prima facie* case without a finding that it was actually tainted. Under the *Coates* rule, an analysis that suffers from major flaws on its face is enough to support a judgment against the defendant in the absence of rebuttal. However, if the statistical analysis was insufficient to establish a *prima facie* case where no allegation of discrimination in the challenged factor had been made—which the *Coates* court apparently would have held—there is no basis for a rule that the *allegation* of discrimination (or weak evidence of discrimination) should have the procedural consequence of shifting the burden of production to the defendant, thereby requiring it to conduct its own statistical analysis by controlling for the important missing variable.

Placing a heavy burden on the defendant to disprove taint is a way of whipsawing the defendant and makes a defendant’s attempt to cope with statistical evidence like boxing with cotton candy. A factual setting similar to that in *Presseisen* will illustrate. Assume that a university that does not discriminate in promotion or in salary has a work force where the most productive scholars are men and that it has a merit system for both promotion and salary.²¹⁸ One would expect that, on average, men would be promoted faster and be paid higher salaries than women, and a regression analysis would show positive, and statistically significant, coefficients for sex both in salary and in time to promotion. Female faculty bring an action for discrimination in both salary and promotion. In defending against the salary discrimination claim, the university attempts to control for rank; after all, the stars who get the fast promotions are the same ones who get the high salaries, and higher rank brings higher salary. The plaintiffs assert that the defendant cannot use the rank variable because there is a statistically significant difference in promotion statistics for men and

218. Because of the requirement of “objective” variables, issues of quality of publications as opposed to quantity would probably be beyond the scope of the regression.

women; therefore, according to the plaintiffs, the variable is tainted. Unless the defendant can affirmatively prove absence of discrimination in promotions—not merely rebut the *prima facie* case—it loses the salary discrimination case. Then, the plaintiffs, having demonstrated salary discrimination, are most of the way toward demonstrating discrimination in promotion, since courts will infer discrimination in one kind of decision based upon proof of discrimination in another.²¹⁹ The non-discriminating employer loses even though it is not discriminating in either salary or promotions and even though the plaintiff has no *good* evidence that it is.

The “tainted variable” issue should seldom arise. The fact that it does is a testament to the sloppiness of thought that is encouraged by current attitudes toward statistical proof. If plaintiffs bring a salary discrimination case and the central issue is whether the variable of performance evaluations can be excluded as tainted, in all likelihood the case should not be a salary discrimination case at all, but rather, if anything, a discriminatory evaluation case. Both sides likely agree that if performance evaluations are controlled for, the salary disparities are eliminated or at least reduced, otherwise they would not be arguing about the variable in the first place. If those with equivalent performance evaluations receive equivalent wages, whether or not there is discrimination in performance evaluations there is no evidence of discrimination in wages. However, treating the case as one for salary discrimination provides plaintiffs the kind of procedural advantages described above if a court is willing to say that the results of performance evaluations cannot be used unless the defendant proves that they are not tainted by discrimination. If the plaintiff characterized the case as what it really is—a discriminatory evaluation case—everyone would agree that the plaintiff bears the burden of proving discrimination in evaluations. Characterizing it as a salary discrimination case allows the court to shift the burden of proving nondiscrimination in evaluations to the defendant.²²⁰

219. *Cf. Cook v. Boorstin*, 763 F.2d 1462, 1468–69 (D.C. Cir. 1985) (“It strikes us as altogether obvious that statistical (or anecdotal) evidence that the Library discriminated against its black librarians would be relevant to whether the Library discriminated against its black attorneys.”); *Vuyanich v. Republic Nat’l Bank*, 505 F. Supp. 224, 338–42 (N.D. Tex. 1980) (holding that discrimination in promotions can be inferred from statistical evidence of discrimination in hiring and salary), *vacated*, 723 F.2d 1195 (5th Cir.), *cert. denied*, 469 U.S. 1073 (1984).

220. In a recent article, the central thrust of which is the sensible proposition that courts should require that statistical models be consistent with economic theory and legal doctrine, James McKeown unfortunately demonstrated the seductiveness of the siren call of tainted variables. McKeown, *supra* note 136, at 654–55 n.86. He suggested that in a sex discrimination action against a university the variable of “publications” in a salary regression would be

5. *The Inexorable Zero*

Another way in which courts have improperly employed statistical proof is in their treatment of circumstances in which the observed number of protected class members is, or approaches, zero.²²¹ In *Teamsters*, the Court rejected the defendant's criticisms of the government's statistical case, stating that "fine tuning of the statistics could not have obscured the glaring absence of minority line drivers" and that the defendant's "inability to rebut the inference of discrimination came not from a misuse of statistics but from 'the inexorable zero.'"²²² Some courts have seized on the "inexorable zero" language to give special treatment to disparities in those cases where there are no women or minorities in a particular job.²²³ For example, in *Capaci v. Katz & Besthoff, Inc.*,²²⁴ the defendant argued that the fact that there were no women in a particular position raised no inference of discrimination because the disparities, if segregated by year, were not statistically significant.²²⁵ The court rejected the defendant's argu-

"tainted"—and therefore should be excluded—if it is more difficult for women to have articles accepted for publication than it is for men. The difficulty with this reasoning is that exclusion in these circumstances would have the effect of holding an innocent employer liable for the acts of the actual discriminator (the discriminatory publications). Thus, unlike a situation where a plaintiff converts what should be a promotion discrimination case against his employer into a salary discrimination case against the same defendant, McKeown's approach would allow what should be a discrimination in publication case to be converted into a salary discrimination case against a completely different defendant. See *Coser v. Moore*, 739 F.2d 746, 750 (2d Cir. 1984):

We also recognize the possibility that the sex characteristics of a particular availability pool may differ from those of the general population because of prior discrimination by entities other than [this defendant, but] where a statistical disparity between men and women may be explained by taking job-related criteria such as prior work experience or prior academic rank into account, evidence of the disparity does not prove unlawful discrimination.

221. See *Boardman & Vining*, *supra* note 49, at 203 ("Total exclusion is sufficient to shift the burden of proof, but such an obvious degree of discrimination is not always necessary.").

222. *International Bhd. of Teamsters v. United States*, 431 U.S. 324, 342 n.23 (1977).

223. See *EEOC v. Andrew Corp.*, 49 Fair Empl. Prac. Cas. (BNA) 804, 815 (N.D. Ill. 1989) ("Since [*Teamsters*], the courts, when confronted with Title VII defendants who have employed or promoted zero or near zero minorities or women, have avoided permitting labor market and statistical analyses to obscure the discrimination inherent in the 'inexorable zero.'"); *Bentley v. City of Thomaston*, 32 Fair Empl. Prac. Cas. (BNA) 1476, 1479 (M.D. Ga. 1983) ("Fine distinctions about whether two or three standard deviations is sufficient to imply discriminatory motive is uncalled for since this court faces the 'inexorable zero' . . ."). Cf. *EEOC v. National Broadcasting Co.*, 753 F. Supp. 452 (S.D.N.Y. 1990), *aff'd mem.*, 940 F.2d 648 (2d Cir. 1991):

Although the fact that no women were hired as Sports Director gives rise to an inference of discrimination—"the inexorable zero" as described in [*Teamsters*]—that inference is weak in light of the very limited number of openings for Sports Director positions. Without a showing that at least one qualified woman applied for a Director position and was denied, EEOC cannot establish a *prima facie* case.

Id. at 466–67.

224. 711 F.2d 647 (5th Cir. 1983), *cert. denied*, 466 U.S. 927 (1984).

225. *Id.* at 654.

ment that "zero is just a number," stating that "it carries special significance in discerning firm policies and attitudes," especially in light of the fact that the disparities were highly significant when individual years were aggregated.²²⁶ However, the defendant was right; a statistically insignificant disparity is no more significant simply because the number is zero.²²⁷ As to revealing something about "firm policies and attitudes," the most that can be said is that it may reveal a firm policy against engaging in affirmative action to avoid "inexorable zeros." Given the absence of any legal requirement to engage in affirmative action, however, that revelation is not a basis for imposing liability.

The district court in *EEOC v. O&G Spring & Wire Forms Specialty Co.*²²⁸ similarly misunderstood the meaning of the number zero. In that case, the EEOC sued the defendant for race and age discrimination. The defendant was a small company owned by a Polish immigrant that hired large numbers of Poles and Hispanics but for the years 1979-1985 had hired no blacks. Defendant's expert testified that the relevant labor market for the defendant was composed disproportionately of recent immigrants who did not speak English, because the defendant needed employees with some skills who were willing to work for low pay. Although the court credited the defendant's testimony, it concluded that "no explanation is sufficient to overcome the

226. *Id.* at 662; see also *Carroll v. Sears, Roebuck & Co.*, 708 F.2d 183, 193 (5th Cir. 1983) (suggesting that the "inexorable zero" adds special weight to the statistical showing); *Valentino v. United States Postal Serv.*, 674 F.2d 56, 73 (D.C. Cir. 1982) ("The 'inexorable zero' . . . can raise an inference of discrimination even if the subgroup analyzed is relatively small.") (citations omitted); *Grant v. Bethlehem Steel Corp.*, 635 F.2d 1007 (2d Cir. 1980), *cert. denied*, 452 U.S. 940 (1981):

Despite evidence of some weaknesses in the statistics, where they disclose a glaring absence of minority representation in the jobs at issue, the burden on the employer increases since "fine tuning" of the statistics will not rebut an inference of discrimination derived "not from a misuse of statistics but from 'the inexorable zero.'"

Id. at 1015 (quoting *Teamsters*, 431 U.S. at 342 n.23); *Wilkins v. University of Houston*, 654 F.2d 388, 410 (5th Cir. 1981) ("Our innate capacity in [statistical] matters extends to 'the inexorable zero' and perhaps, unevenly, somewhat beyond; but the day is long past—past at least since the Supreme Court's sophisticated analysis in *Castaneda v. Partida* . . .—when we proceed with any confidence toward broad conclusions from crude and incomplete statistics."), *vacated*, 459 U.S. 809 (1982).

227. See *Kaye*, *supra* note 33, at 1345 n.59 ("Even the 'inexorable zero,' which the courts took to be dramatic evidence of discrimination in the days before hypothesis testing in court, may not be sufficient to warrant rejection of the null hypothesis at the .05 level."). But see *Elaine W. Shoben, The Use of Statistics to Prove Intentional Employment Discrimination*, LAW & CONTEMP. PROBS., Autumn 1983, at 221, 238 ("The reason that the inexorable zero is so compelling is that such a result is so unlikely to happen without the impermissible influence in the decisionmaking.").

228. 705 F. Supp. 400 (N.D. Ill. 1988).

'inexorable zero' employment of blacks at O & G from 1979 through 1985."²²⁹

Courts' treatment of the "inexorable zero" is simply an act of statistical legerdemain—a way of holding defendants liable on the basis of statistical evidence even when such findings are inappropriate under the prevailing rules governing such evidence. In other words, it is a way to hold defendants liable on the basis of evidence that otherwise would not warrant a finding of discrimination. Even if statistically significant disparities were strongly probative of discrimination, statistically insignificant disparities are not, even if the representation is zero.

IV. PRESCRIPTION

Acknowledgment of the Statistical Fallacy compels a recognition that statistical evidence, standing alone, is only weakly probative on the question of whether an observed disparity is due to chance or nonchance factors. Moreover, even when one may confidently conclude that a disparity is so large that it must have been caused by nonchance factors, statistical inference provides no insight at all into whether the nonchance factor was an impermissible one, especially where the statistical analysis contains only rudimentary control for relevant variables. Consequently, courts should be substantially more demanding in the sophistication required of plaintiffs' statistical showings and extremely wary of concluding that discrimination has occurred on the basis of a largely statistical case. Further, in evaluating the statistical evidence, courts should not shift the burden of proof to defendants under the guise of evidentiary rules.

A. *The Need for Strong Anecdotal Evidence*

Because statistically significant disparities by themselves are at best only weakly probative of discrimination, a plaintiff should be obliged to demonstrate substantially more than a statistical disparity in order

229. *Id.* at 406; see also *EEOC v. Andrew Corp.*, 49 Fair Empl. Prac. Cas. (BNA) 804, 816 (N.D. Ill. 1989) ("The data in this case respecting Black office and clerical workers is dominated by the 'inexorable zero' and cannot be 'explained away.'").

The employer in *O&G Spring* had no applications for the years prior to 1984, and the court found that the earlier applications had been "legally destroyed." *O&G Spring*, 705 F. Supp. at 402-03. For the years 1984 and 1985, the company had received four applications from blacks out of a total of 58. *Id.* at 403. The court also had before it testimony that the racial composition of the neighborhood was rapidly changing; it had been 10-15 percent black in 1984 and approximately 50 percent black in 1986. *Id.* at 405. This suggests that extrapolating backward from the black representation of the applicant pool from 1984-85 to reconstruct the applicant flow from 1979 to 1983 would be highly questionable.

to establish by a preponderance of the evidence that discrimination is the employer's standard operating procedure. If the employer has in fact routinely engaged in discrimination, then by definition individual victims abound, and the plaintiff ought in fairness to bear the burden of providing evidence of a substantial number of individual cases of discrimination. Many courts already expect to see some "anecdotal" evidence of discrimination—that is, testimony concerning individual instances of discrimination.²³⁰ However, a number of courts have held that anecdotal evidence of discrimination is not required, especially if the statistical evidence is deemed strong.²³¹ Where the statistical evidence is weak, courts have tended to demand stronger anecdotal evidence²³² but have stopped far short of requiring a quantity of evidence sufficient to buttress a case grounded on weak statistical proof.

230. See, e.g., *Sheehan v. Purolator Courier Corp.*, 839 F.2d 99, 103 (2d Cir.) (holding that the district court "was not clearly erroneous in concluding . . . that submitting an affidavit from only one aggrieved employee, other than the named plaintiffs, was insufficient to establish a class of aggrieved individuals"), *cert. denied*, 488 U.S. 891 (1988); *Rossini v. Ogilvy & Mather, Inc.*, 798 F.2d 590, 604 (2d Cir. 1986) ("In evaluating all of the evidence in a discrimination case, a district court may properly consider the quality of any anecdotal evidence or the absence of such evidence."); *Coates v. Johnson & Johnson*, 756 F.2d 524, 532 (7th Cir. 1985) ("The plaintiffs' prima facie case will thus usually consist of statistical evidence demonstrating substantial disparities in the application of employment actions as to minorities and the unprotected group, buttressed by evidence of general policies or specific instances of discrimination."); *Coser v. Moore*, 739 F.2d 746, 752 (2d Cir. 1984) (holding that the "failure to locate and identify a meaningful number of concrete examples of discrimination" constitutes major weakness in employment discrimination case); *Goff v. Continental Oil Co.*, 678 F.2d 593, 597 (5th Cir. 1982) ("[E]ven if all three witnesses' accounts of racial discrimination were true, this evidence would not have been enough to prove a pattern or practice of company-wide discrimination by Conoco."); *Ste. Marie v. Eastern R.R. Ass'n*, 650 F.2d 395, 405-07 (2d Cir. 1981) (holding that where relevant statistical evidence was lacking, seven individual incidents of discrimination were insufficient to demonstrate pattern or practice of discrimination).

231. In *Hazelwood School District v. United States*, 433 U.S. 299 (1977), the Court stated that "[w]here gross statistical disparities can be shown, they alone may in a proper case constitute prima facie proof of a pattern or practice of discrimination." *Id.* at 307-08. Similarly, in *Segar v. Smith*, 738 F.2d 1249 (D.C. Cir. 1984), *cert. denied sub nom. Meese v. Segar*, 471 U.S. 1115 (1985), the court held: "when a plaintiff's statistical methodology focuses on the appropriate labor pool and generates evidence of discrimination at a statistically significant level, no sound policy reason exists for subjecting the plaintiff to the additional requirement of either providing anecdotal evidence or showing gross disparities." *Id.* at 1278; see also *Catlett v. Missouri Highway Transp. Comm'n*, 828 F.2d 1260, 1265 (8th Cir. 1987) (stating that either statistical evidence or anecdotal evidence alone may be sufficient to establish a pattern or practice of discrimination), *cert. denied*, 485 U.S. 1021 (1988); *Coates*, 756 F.2d at 533 ("Neither statistical nor anecdotal evidence is automatically entitled to reverence to the exclusion of the other.").

232. See, e.g., *EEOC v. Sears, Roebuck & Co.*, 839 F.2d 302, 311 (7th Cir. 1988) ("[E]xamples of individual discrimination are not always required, but we think that the lack of such proof reinforces the doubt arising from the questions about validity of the statistical evidence.") (quoting *Griffin v. Board of Regents of Regency Univs.*, 795 F.2d 1281, 1292 (7th Cir. 1986)); *Woodbury v. New York City Transit Auth.*, 832 F.2d 764, 771 (2d Cir. 1987)

The philosophy behind allowing plaintiffs to prevail on largely statistical evidence is that employers can surreptitiously discriminate with substantial impunity and that statistical evidence is a necessary means for smoking out employers who otherwise could not be caught.²³³ That justification cannot withstand analysis. First, it should be clear from the foregoing discussion that although a plaintiff may "prove" a case with statistical evidence, it is doubtful that there is much correlation between plaintiff victories that rest on such proof and defendant culpability. Put another way, while statistics may allow a plaintiff to win who would otherwise not win, this result cannot argue in favor of statistical proof unless the plaintiff actually deserved to win under the applicable substantive law. But that, of course, is a matter that we can scarcely assume, since the plaintiff was spared the need to establish such a case. Second, the assumption that an employer can engage in widespread and systematic discrimination and leave no sign of it other than gross statistical patterns defies belief. Since 1965, thousands of successful individual disparate-treatment actions have been brought, demonstrating that such cases are not foredoomed to failure. If discrimination is truly widespread, even the relatively slight rebuttal burden under *McDonnell Douglas* will assume massive proportions when the employer is forced to articulate large numbers of pretextual, yet internally consistent, nondiscriminatory explanations. It must be remembered that the employer cannot simply put forward *ad hoc* explanations for each discriminatory decision; it must make sure that its justification for one action is not inconsistent with its justification for another. If the plaintiffs in a pattern-or-practice case cannot find evidence of large numbers of individual discriminatory acts, it is quite likely because there is no evidence to find.

(holding that anecdotal evidence consisting of the way in which the defendant resolved a disciplinary matter involving a white officer and "the 'sincere impression of witness after witness' that the 'disciplinary system deals unequally with the races' " insufficient "given the complete absence of creditable statistical evidence"); *Valentino v. United States Postal Serv.*, 674 F.2d 56, 69 (D.C. Cir. 1982) ("[W]hen the statistical evidence does not adequately account for 'the diverse and specialized qualifications necessary for [the positions in question],' strong evidence of individual instances of discrimination becomes vital to the plaintiff's case.") (quoting *Wilkins v. University of Houston*, 604 F.2d 388, 410 (5th Cir. 1981)) (alteration in original); *Garcia v. Rush-Presbyterian-St. Luke's Medical Ctr.*, 660 F.2d 1217, 1225 (7th Cir. 1981) ("We find very damaging to plaintiffs' position the fact that not only was their statistical evidence insufficient, but that they failed completely to come forward with any direct or anecdotal evidence of discriminatory employment practices by defendants. Plaintiffs did not present in evidence even one specific instance of discrimination."). Cf. *Rossini*, 798 F.2d at 604 (quoting district court for proposition that "strong statistical proof . . . is critical [in the instant case], since there is no anecdotal evidence supporting the figures.") (footnote omitted) (alteration in original).

233. See *International Bhd. of Teamsters v. United States*, 431 U.S. 324, 339 n.20 (1977); *Frazier v. Consolidated Rail Corp.*, 851 F.2d 1447, 1453 (D.C. Cir. 1988).

In order for anecdotal evidence to provide the quantum of logical support needed to buttress statistical showings—which, as we have seen, are far less consequential than previously recognized—courts should require substantial quantities of such evidence.²³⁴ In large class cases, tens of thousands of employment decisions may be at issue. The fact that plaintiffs can establish by a preponderance of the evidence that one or two dozen discriminatory employment decisions were made during the liability period, which in a class action is often many years, is not strong evidence that the employer has engaged in a pattern or practice of discrimination.²³⁵ To establish a pattern or practice requires more than “the mere occurrence of isolated or ‘accidental’ or sporadic discriminatory acts.”²³⁶ As the Supreme Court has recognized, “a piece of fruit may well be bruised without being rotten to the core.”²³⁷ It is doubtful that those who believe that a few instances of individual discrimination should be taken to be representative of the employer’s standard operating procedure would accept the converse proposition that an employer could rebut a plaintiff’s statistical evidence by calling a few minority employees to testify that they had not been discriminated against.²³⁸

234. See *Chisholm v. United States Postal Serv.*, 665 F.2d 482, 495 (4th Cir. 1981) (upholding finding of discrimination where anecdotal evidence came from 20 class members).

235. See *O'Donnell Constr. Co. v. District of Columbia*, 963 F.2d 420, 427 (D.C. Cir. 1992) (“[W]hile anecdotal evidence may suffice to prove individual claims of discrimination, rarely, if ever, can such evidence show a systemic pattern of discrimination.” . . . Anecdotal evidence is most useful as a supplement to strong statistical evidence—which the Council did not produce in this case.”) (citation omitted); see also *General Tel. Co. v. Falcon*, 457 U.S. 147 (1982):

Even though evidence that [plaintiff] was passed over for promotion when several less deserving whites were advanced may support the conclusion that respondent was denied the promotion because of his national origin, such evidence would not necessarily justify the additional inferenc[e] . . . that this discriminatory treatment is typical of petitioner’s promotion practices

Id. at 158; *Rocha v. Great Am. Ins. Co.*, 850 F.2d 1095, 1102 (6th Cir. 1988) (Guy, J., dissenting) (arguing that anecdotal evidence regarding only five out of defendant’s 3,200 employees was insufficient to establish a pattern or practice of age discrimination).

236. *Teamsters*, 431 U.S. at 336; see also *King v. General Elec. Co.*, 960 F.2d 617, 623 (7th Cir. 1992) (“In order to prove that the employer has engaged in a pattern or practice of discrimination, then, the plaintiff must show that there is regular, purposeful, less-favorable treatment of a protected group.”).

237. *Cooper v. Federal Reserve Bank*, 467 U.S. 867, 880 (1984).

238. One form of anecdotal evidence that should be strictly barred is noncompliance with an affirmative-action plan. Courts have taken differing approaches to this issue. Compare *Gonzales v. Police Dep’t*, 901 F.2d 758, 760–61 (9th Cir. 1990) (holding that the district court erred in failing to consider as evidence of pretext in an individual disparate treatment case the employer’s failure to live up to its affirmative-action plan that had been adopted as part of a consent decree) and *Craik v. Minnesota State Univ. Bd.*, 731 F.2d 465, 472 (8th Cir. 1984) (Although “[n]either Title VII, 42 U.S.C. § 1983, nor the Fourteenth Amendment requires an employer to institute an affirmative-action program, . . . evidence that an employer has failed to live up to an affirmative

In *EEOC v. Sears, Roebuck & Co.*,²³⁹ the EEOC attempted to turn the weak probative value of anecdotal evidence to its advantage. Arguing that its failure to produce anecdotal evidence should not be deemed to undermine its statistical evidence, the EEOC suggested that introduction of such evidence would be “‘inappropriate’ because ‘where 47,000 hires and promotions were at issue . . . it would have been impossible to present enough individual demonstrations [sic] of discrimination to meaningfully reflect on the statistics.’”²⁴⁰ In a sense, of course, the EEOC was right. Even if it presented evidence that forty of the 47,000 hiring and promotion decisions were probably

action plan is relevant to the question of discriminatory intent.”) and *Taylor v. Teletype Corp.*, 648 F.2d 1129, 1135 n.14 (8th Cir.) (stating that evidence of the defendant’s failure to live up to its affirmative action plan is “relevant to discerning the Company’s attitude regarding race”), *cert. denied*, 454 U.S. 969 (1981) and *Morman v. John Hancock Mut. Life Ins. Co.*, 672 F. Supp. 993, 995 (E.D. Mich. 1987) (“For policy reasons implicit in our antidiscrimination laws, [a]n employer cannot be allowed to advertise itself as an affirmative action employer and then wink at its [affirmative-action plan] when it gets in the way.”) (quoting *Fang-Hui Liao v. Dean*, 658 F. Supp. 1554, 1561 (N.D. Ala. 1987)) (first alteration in original) with *Liao v. Tennessee Valley Auth.*, 867 F.2d 1366, 1369 (11th Cir. 1989) (“[T]he failure to give a preference under such a plan cannot be used to support an allegation of discrimination in employment decisions.”), *cert. denied*, 494 U.S. 1078 (1990) and *Manoharan v. Columbia Univ.*, 842 F.2d 590, 594 (2d Cir. 1988) (“Since section [703(j)] of Title VII states that affirmative action in employment practices is not required by Title VII . . . an employer’s failure to follow its own voluntary affirmative action program cannot, by itself, constitute an unlawful employment practice”) and *Ferguson v. Veterans Admin.*, 723 F.2d 871, 873 (11th Cir.) (“[T]o the extent plaintiff’s claim is based on the failure of her employer to follow an affirmative action plan . . . she loses as a matter of law, absent a showing of discriminatory implementation of that plan.”), *cert. denied*, 469 U.S. 1072 (1984).

For a whole host of reasons, failure to follow an affirmative action plan should not enhance a plaintiff’s case. Under Title VII, an employer is not obligated to have an affirmative action plan at all. Section 703(j) provides:

Nothing contained in this subchapter shall be interpreted to require any employer . . . to grant preferential treatment to any individual or to any group because of the race [or] sex . . . of such individual or group on account of an imbalance which may exist with respect to the total number or percentage of persons of any race [or] sex . . . employed by any employer

42 U.S.C. § 2000e-2(j) (1988). Holding an employer liable because of inadequate affirmative action, whether inadequacy is measured against the employer’s affirmative action plan or a judge’s notion of proper affirmative action efforts, is to hold it liable precisely because of its alleged failure “to grant preferential treatment . . . on account of an imbalance.”

The fact that an employer has a voluntary plan that it has failed to follow in all respects—whether through ineptness or by design—does not logically suggest an invidious intent to discriminate against women and minorities. Moreover, as a policy matter, if voluntary affirmative action is desirable, treating a failure to comply with a plan as evidence of illegal conduct creates a substantial incentive for an employer to forego affirmative action altogether or to construct a plan so vague that it could never be said to have violated it. Thus, penalizing employers for unsuccessful or unenthusiastic affirmative-action efforts creates perverse incentives.

239. 839 F.2d 302 (7th Cir. 1988).

240. *Id.* at 311 (bracketed material in original) (citation omitted). The trial involved 20,000 pages of transcripts, 49 witnesses, and 2172 exhibits totaling over 22,000 pages. *Id.* at 307 n.2.

discriminatory, that proof would relate to less than one-tenth of one percent of the total. Because these forty cases would have been carefully selected by the EEOC and could hardly be considered a representative sample of Sears' employment decisions, their probative value on the ultimate question—whether discrimination was Sears' standard operating procedure—would be little more powerful than identifying forty black criminals and arguing from that sample that blacks have a propensity toward crime.

The EEOC was correct in its essential position that statistical evidence of discrimination is not made substantially stronger by weak anecdotal evidence. The important recognition that even "statistically significant" disparities, standing alone, are very weak evidence of intentional discrimination would be undermined if a court were persuaded to impose class-wide liability on the basis of a few anecdotal accounts. Although upholding a judgment for the defendant, the Seventh Circuit in *Sears* appeared to accord more value to small amounts of anecdotal evidence than is warranted. It suggested that examples of individual instances of discrimination need not be numerous and that "[e]ven a few examples would have helped bring 'cold numbers convincingly to life.'"²⁴¹ The quoted phrase, which comes from the Supreme Court's *Teamsters* decision, is commonly invoked, but without much thought about why a few, perhaps isolated, instances of discrimination are to be taken as strong evidence that statistical disparities are caused by systematic discrimination.

In *Segar v. Smith*,²⁴² the D.C. Circuit was even less concerned than the *Sears* court with anecdotal evidence. In *Segar*, the plaintiffs had introduced statistical evidence that was considered to be strong and anecdotal accounts of specific instances of discrimination that the district court had refused to credit.²⁴³ Rather than concluding that the plaintiffs' inability to find any flesh-and-blood victims of discrimination suggested a weakness in the statistics, the court suggested that requiring anecdotal evidence "would reflect little more than a superstitious hostility to statistical proof, a preference for the intuitionistic and individualistic over the scientific and systemic."²⁴⁴ However, as Douglas Laycock has noted, the *Segar* court's "blind faith in statistical

241. *Id.* at 311–12; cf. *EEOC v. Chicago Miniature Lamp Works*, 947 F.2d 292, 303 (7th Cir. 1991) (suggesting that "[t]he EEOC might have overcome deficiencies in its statistics if it had presented more compelling anecdotal evidence").

242. 738 F.2d 1249 (D.C. Cir. 1984), *cert. denied sub nom. Meese v. Segar*, 471 U.S. 1115 (1985).

243. *Id.* at 1264 n.10.

244. *Id.* at 1278.

evidence without consideration of the assumptions underlying the statistical techniques reflects a superstitious faith in the *pseudoscientific*.”²⁴⁵ If one observes statistical disparities that might suggest discrimination against blacks but then determines by looking at individual cases that what initially appeared discriminatory from the statistics has a nondiscriminatory explanation, one ought to question whether one is drawing the correct inference from the statistics.²⁴⁶

In *EEOC v. American National Bank*,²⁴⁷ the Fourth Circuit elevated faith in statistics to an even higher level than the *Segar* court. The district court had found that the EEOC had established a *prima facie* case through static work-force statistics—that is, a comparison of the composition of the employer’s work force to the composition of the relevant labor market—but it also found that the defendant had rebutted what the court viewed as a relatively weak *prima facie* case²⁴⁸ through the use of applicant-flow statistics, demonstrating that within the liability period there had been no significant shortfall in black hiring.²⁴⁹ The district court also considered the EEOC’s presentation of testimony from thirty-one alleged individual victims of discrimination and concluded that not one of them had been subjected to discrimination.²⁵⁰ Because the EEOC had stipulated that there was a maximum of fifty-two potential victims of discrimination,²⁵¹ only twenty-one other potential victims of discrimination existed. The court found no evidence that any of these persons had been subjected to discrimina-

245. Douglas Laycock, *Statistical Proof and Theories of Discrimination*, LAW & CONTEMP. PROBS., Autumn 1986, at 97, 103 (emphasis added).

246. Cf. *Catlett v. Missouri Highway & Transp. Comm’n*, 828 F.2d 1260, 1266 (8th Cir. 1987) (upholding a finding of class-wide sex discrimination in highway-maintenance positions, notwithstanding jury verdicts in favor of the defendant on the individual claims of all four class representatives), *cert. denied*, 485 U.S. 1021 (1988); *Boykin v. Georgia-Pacific Corp.*, 706 F.2d 1384, 1393 (5th Cir. 1983) (stating that the fact that the district court finds that class representatives and other testifying employees were not subjected to discrimination does not invalidate the class claim), *cert. denied*, 465 U.S. 1006 (1984).

247. 652 F.2d 1176 (4th Cir. 1981), *cert. denied*, 459 U.S. 923 (1982).

248. Disparities were in the one to three standard deviation range.

249. *American Nat’l Bank*, 652 F.2d at 1181.

250. *EEOC v. American Nat’l Bank*, 21 Fair Empl. Prac. Cas. (BNA) 1532, 1584 (E.D. Va. 1979), *aff’d in part, rev’d in part, and vacated in part*, 652 F.2d 1176 (4th Cir. 1981), *cert. denied*, 459 U.S. 923 (1982).

251. *Id.* at 1569 (51 potential victims in addition to the original claimant). The EEOC and the bank had stipulated that “the only [possible] potential discriminatees” were all qualified black applicants who applied for a position in the six-month period prior to the hiring of a white, because the bank had an unchallenged policy of considering only applications filed within six months of the creation of a vacancy. *American Nat’l Bank*, 652 F.2d at 1213 (Russell, J., dissenting.)

tion.²⁵² The district court believed so strongly in the weakness of the EEOC's case that it concluded that the case was "unreasonable, vexatious, and litigated in bad faith"²⁵³ and awarded attorneys fees to the defendant.

Notwithstanding the fact that the district court had made specific findings that none of the "potential discriminatees" had been a victim of discrimination, the court of appeals reversed the district court's judgment in favor of the defendant and ruled that the EEOC was entitled to judgment *as a matter of law*.²⁵⁴ It reasoned that the statistical applicant-flow data was not entirely complete and therefore could not rebut the prima facie case established by the static work-force comparison. The court also ruled, without providing a rationale, that the individual evidence could not rebut the prima facie case.²⁵⁵ Thus, even in the face of the district court's specific findings that none of the potential victims of discrimination had in fact been discriminated against—findings that the Fourth Circuit did not disturb on appeal—the court deemed the statistical proof so powerful as to compel a judgment in favor of the EEOC. Such a conclusion is possible only through suspension of common sense.²⁵⁶

252. *American Nat'l Bank*, 21 Fair Empl. Prac. Cas. (BNA) at 1592. In ruling on the merits of the case, the court declined to consider evidence concerning these persons, because the EEOC stipulated that its case would be limited to those claimants who testified at trial. *Id.* at 1585. However, in ruling on the defendant's motion for attorneys fees, the court considered whether there was a possibility that any of the 21 had been discriminated against on the basis of race. *Id.* at 1584–85.

253. *Id.* at 1593.

254. *American Nat'l Bank*, 652 F.2d at 1181.

255. *Id.* at 1200; cf. *Coates v. Johnson & Johnson*, 756 F.2d 524, 532 (7th Cir. 1985) ("The pattern or practice claim may also fail—despite any statistical evidence offered by plaintiffs—if the defendant articulates a nondiscriminatory, nonpretextual reason for every discharge."); *Paxton v. Union Nat'l Bank*, 688 F.2d 552, 567 (8th Cir. 1982), *cert. denied*, 460 U.S. 1083 (1983). One of the problems faced by defendants in attempting to provide applicant-flow data is that if records of the race of applicants are maintained, some courts have inferred that the racial data are being collected for an improper purpose. For example, the district court in *EEOC v. Chicago Miniature Lamp Works*, 947 F.2d 292, 304 (7th Cir. 1991), seems to have relied upon the fact that applications from blacks had the letter "B" written by hand on them as anecdotal evidence of discrimination: "The court did not specifically find that this was invidious 'race-coding,' but instead obliquely cited another case in which that conclusion had been made. The EEOC concedes in its brief that the labeling of the applications 'could have had some legitimate purpose.'"

256. See *American Nat'l Bank*, 652 F.2d at 1220 (Russell, J., dissenting) (footnote omitted):

[S]tatistics, with their inferences, cannot stand against proof that an employer never engaged in discrimination in a single employment decision in the relevant time period, when reviewed in the light of the available applicant flow. This positive proof that no employment decision was tainted is 'the proof of the pudding' in the establishment of no discrimination and it completely nullifies any inference sought to be drawn from statistical evidence.

In sum, evidence of actual instances of discrimination should be required in all cases because statistical evidence by itself proves little. Moreover, evidence of a substantial number of victims is necessary in order to ensure that the individual discriminatory actions are representative of the employer's regular practices and not merely isolated instances. Rather than anecdotal evidence being viewed as an adjunct to statistical evidence, anecdotal evidence should be the core of the case, demonstrating large numbers of individual instances of discrimination. The function of statistical evidence should be simply to support the inference that those many instances are indeed part of a larger pattern or to attempt some estimate of the total magnitude of the discrimination.²⁵⁷

B. The Need for "Gross Disparities"

In *Teamsters*, the Supreme Court stated that "[e]vidence of *longlasting and gross disparity* between the composition of a work force and that of the general population" may be evidence of intentional discrimination.²⁵⁸ Similarly, in *Hazelwood*, the Court stated that "[w]here *gross statistical disparities* can be shown, they alone may in a proper case constitute prima facie proof of a pattern or practice of discrimination."²⁵⁹ Some subsequent lower-court cases, however, have expressly

See also Barnes v. GenCorp, Inc., 896 F.2d 1457, 1469 (6th Cir.) (holding where the employer presents evidence that each plaintiff is less qualified than others occupying comparable positions, defendant undercuts plaintiffs' statistical proof; unless plaintiffs can demonstrate that explanation is pretextual, employer entitled to summary judgment), *cert. denied*, 111 S. Ct. 211 (1990). *But see* Coates, 756 F.2d at 533 ("[T]he class claim does not fail just because the district court finds that the company has satisfactorily explained the discharges of the named class representatives and any other testifying employees"; "a defendant's successful rebuttal of each alleged instance of discrimination weakens, but does not defeat, a plaintiff's class claim."); Boykin v. Georgia-Pacific Corp., 706 F.2d 1384, 1393 (5th Cir. 1983) (holding that the defendants cannot combat a case of class-wide discrimination with evidence about a few of the promotion decisions), *cert. denied*, 465 U.S. 1006 (1984).

257. In *Domingo v. New England Fish Co.*, 727 F.2d 1429 (9th Cir. 1984), the court held that the district court's reliance on a statistical comparison that failed to take into account the number of minorities qualified to perform the jobs at issue was not reversible error where there was sufficient nonstatistical evidence of disparate treatment. According to the court, the statistics "merely demonstrated the consequences of [defendant's] discriminatory hiring practices." *Id.* at 1436. Because in a proper case a pattern or practice of discrimination can be established by nonstatistical evidence alone—for example, by introducing a large amount of anecdotal evidence or by establishing a formal policy of discrimination—the court may have been correct that reliance on the statistics was not reversible error. However, the Ninth Circuit was wrong to the extent that it was suggesting that statistical evidence that did not account for qualifications "demonstrated" anything.

258. *International Bhd. of Teamsters v. United States*, 431 U.S. 324, 339–40 n.20 (1977) (emphasis added).

259. *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 307–08 (1977) (emphasis added).

denied that a statistical disparity need be "gross," as long as it is statistically significant.²⁶⁰

Allowing a finding of liability to be based upon disparities that are statistically significant but not "gross" places undue faith in statistical models and threatens to result in improper findings of liability. No one doubts that statistical models are merely approximations of reality—some variables are omitted, others are represented by proxy. While there may be arguments about whether a given approximation is "good enough" to be relevant, no one believes that they present a perfect picture. Because with large samples even trivial differences in selection rates may be statistically significant,²⁶¹ the fact that statistical models are only approximations of the selection process assumes greater importance in cases involving large numbers of employment decisions.

Assume, for example, that for a given entry-level job the interest of men is slightly greater than women. As a result, for every hundred employees, 51 will be men and 49 will be women, despite a parity of representation in the relevant labor pool. If the employer has 100 employees, the underrepresentation of women would be statistically insignificant; in fact, there is only a one in twelve chance that a 50/50 ratio would be obtained even if the true likelihood of selection for men and women were identical.²⁶² If, however, the employer had 10,000 employees (5100 men and 4900 women), the underrepresentation of women would be statistically significant at the 5-percent level.²⁶³ Thus, the incorrect assumption of equal interest has only a trivial

260. See, e.g., *Segar v. Smith*, 738 F.2d 1249, 1278 (D.C. Cir. 1984) ("[W]hen a plaintiff's statistical methodology focuses on the appropriate labor pool and generates evidence of discrimination at a statistically significant level, no sound policy reason exists for subjecting the plaintiff to the additional requirement of . . . showing gross disparities."), *cert. denied sub nom. Meese v. Segar*, 471 U.S. 1115 (1985); see also *Black Shield Police Ass'n v. City of Cleveland*, 42 Fair Empl. Prac. Cas. (BNA) 270, 275 (N.D. Ohio 1986) (When statistical evidence is "finely tuned to the relevant labor pool, gross disparities need not be shown to permit an inference of discrimination.") (quoting *Segar*, 738 F.2d at 1278).

261. See Kaye, *supra* note 33, at 1346.

262. The standard deviation would be:

$$\sqrt{np(1-p)} = \sqrt{(100)(.5) \times (1-.5)} = \sqrt{25} = 5$$

The z-score would be:

$$z = \frac{(\text{observed-expected})}{\text{s.d.}} = \frac{(49-50)}{5} = -.20$$

The associated probability of a deviation this great from the expected is approximately 92 percent.

263. The standard deviation would be:

$$\sqrt{np(1-p)} = \sqrt{(10,000)(.5) \times (1-.5)} = \sqrt{2500} = 50$$

effect in a small work force but a possibly outcome-determinative effect in a large one.²⁶⁴

Because of this large-sample effect, Meier, Sacks and Zabell have argued that the "4/5ths rule of thumb" employed by the EEOC in determining whether an employment practice has a substantial disparate impact should be carried over into pattern-or-practice cases.²⁶⁵ The rule would have a somewhat different function in pattern-or-practice cases, however. The 4/5ths rule in disparate-impact cases has a substantive function: an employment practice that has only a small disparate impact does not have the "substantial disparate impact" that establishes a *prima facie* case; application of the 4/5ths rule ensures that only substantial impacts will result in liability. In intentional discrimination cases, however, the 4/5ths rule would serve an evidentiary function. That is, the rule does not authorize intentional discrimination up to a certain threshold; instead, it simply means that unless the deviation between observed and expected numbers is of a sufficient magnitude, we cannot be confident that there is a real difference at all.²⁶⁶

The crudeness of even the most sophisticated model of the employment process means that extraordinary confidence in statistical models is not warranted and, if statistical evidence is to play any role in discrimination cases, some method of distinguishing between gross and nongross disparities must be adopted. Any boundary between the two types of disparities is necessarily arbitrary, but if one is to be chosen, the 4/5ths rule may be as good as any, although it would, for reasons

The z-score would be:

$$z = \frac{(\text{observed-expected})}{\text{s.d.}} = \frac{(4900-5000)}{50} = -2$$

The associated probability of a deviation this great from the expected is less than 5 percent.

264. In a case like that hypothesized, it is difficult to understand how the defendant might hope to rebut the *prima facie* statistical case.

265. Meier et al., *supra* note 50, at 159–61. Under the 4/5ths rule, an employment practice is deemed to have a disparate impact if the "selection rate for any race, sex, or ethnic group . . . is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate." 29 C.F.R. § 1607.4(D) (1992).

Some commentators have criticized the 4/5ths rule on the ground that it fails to take into account whether disparities are also statistically significant. See Boardman & Vining, *supra* note 49, at 211–17; Braun, *supra* note 17, at 78–81; Elaine W. Shoben, *Differential Pass-Fail Rates in Employment Testing: Statistical Proof Under Title VII*, 91 HARV. L. REV. 793, 805–11 (1978). Meier, Sacks, and Zabell argue that the EEOC's rule, properly understood, does in fact incorporate a test of statistical significance. Meier et al., *supra* note 50, at 166–68.

266. See also Smith & Abram, *supra* note 13, at 53 ("The Court's emphasis in *Teamsters* and in *Hazelwood* on evidence of 'long-lasting and gross disparities' suggests that more is required in the evaluation of statistical proof of disparate impact than statistically significant disparities.").

already discussed, be wiser to forswear the use of statistical evidence except as background data.

C. The Need to Limit Introduction of Evidence Concerning Probabilities

In the fifteen years since the Supreme Court placed its imprimatur on statistical proof of discrimination in Title VII cases, courts have demonstrated themselves to be less than competent in dealing with statistical evidence. The Statistical Fallacy, embraced on a wholesale basis by the courts, has caused them consistently to overestimate the importance of statistically significant disparities in an employer's work force. In large part because of the "suspect" nature of these disparities, courts have inappropriately shifted to the defendant the burden of justifying them.

The incapacity of courts to deal with this form of evidence has been recognized in a peculiar back-handed fashion by appellate courts in the level of deference that they give to lower courts' statistical findings. For example, the Seventh Circuit has in several cases relied upon a rule that "especially where statistical evidence is involved, great deference is due the district court's determination of whether the resultant numbers are sufficiently probative of the ultimate fact in issue."²⁶⁷ Although the court has not identified the basis for its rule, it must be different from the rationale for granting special deference to a trial court's resolution of credibility because of its ability to observe the witnesses' demeanor.²⁶⁸ The Fifth Circuit has more candidly revealed its justification for deference to the district court, stating that "we do not have the statistical expertise to declare a particular statistical technique inappropriate where a qualified expert expresses a contrary view

267. *Soria v. Ozinga Bros., Inc.*, 704 F.2d 990, 994-95 n.6 (7th Cir. 1983); see also *Mozee v. American Commercial Marine Serv. Co.*, 940 F.2d 1036, 1045 (7th Cir. 1991) (quoting *Soria*); *Ottaviani v. State Univ.*, 875 F.2d 365, 375 (2d Cir. 1989) (adopting Seventh Circuit rule), *cert. denied*, 493 U.S. 1021 (1990); *EEOC v. Sears, Roebuck & Co.*, 839 F.2d 302, 310 (7th Cir. 1988) (quoting *Soria*).

268. See *Anderson v. City of Bessemer City*, 470 U.S. 564, 574 (1985) (stating that the "clearly erroneous" standard of Rule 52 of the Federal Rules of Civil Procedure is applicable "even when the district court's findings do not rest on credibility determinations, but are based instead on physical or documentary evidence or inferences from other facts").

In 1985, Rule 52 was amended to make clear that "[f]indings of fact, *whether based upon oral or documentary evidence*, shall not be set aside unless clearly erroneous." FED. R. CIV. P. 52 (emphasis added). The Notes of the Advisory Committee point out that although many courts had given reduced deference to findings not based upon assessments of credibility, interests of "stability and judicial economy"—not the trial court's superior ability to draw inferences from documentary evidence—mandate deference to the trier of fact. *Id.* (Advisory Committee's Note).

and the district court credits the testimony.”²⁶⁹ Thus, it is neither the district court’s superior vantage point nor institutional efficiency that entitles its findings to deference; it is the appellate court’s lack of competence to select between the statistical presentations of competing expert witnesses. But surely the district court’s competence is no greater. If a decision either way can rarely be declared clearly erroneous, neither can it be declared clearly correct or even, perhaps, probably correct.

The increased availability of jury trials under the Civil Rights Act of 1991²⁷⁰ threatens to exacerbate the problem of triers of fact dealing inappropriately with statistical evidence. Judges who today deal routinely with statistical evidence in litigation presumably are better able to evaluate it than jurors for whom a trial may be the first and last time they ever deal with complex statistical analyses. If judges cannot competently evaluate such evidence, there is every reason to believe that juries are even less able.²⁷¹ Moreover, the difficulty of reviewing a jury’s reasoning makes it more difficult to enforce rigor in the fact-finding process.

It must be now recognized that a great deal of caution is necessary in evaluating statistical evidence in discrimination cases. Hypothesis testing, with its reliance on the assumption that the resultant p-value represents the probability that the observed distribution was a consequence of chance and its declaration of results as “statistically significant,” should be abandoned altogether.²⁷² Such evidence is simply irrelevant to the ultimate question. Evidence of the p-value itself should also be excluded, although it is admittedly relevant in some cases. Even though it is not a measure of the probability that a particular employer’s work-force disparity is a product of chance, the p-value is still relevant because it gives a *subjective* intuitive impression

269. *Rendon v. AT&T Technologies*, 883 F.2d 388, 397 (5th Cir. 1989).

270. Pub. L. No. 102-166, 105 Stat. 1071 (codified at 42 U.S.C. §§ 1981, 2000e (1988 & 1992 Supp.)).

271. See *King v. General Elec. Co.*, 960 F.2d 617, 624 (7th Cir. 1992) (suggesting that one of the difficulties of employing Title VII pattern-or-practice analysis in cases brought under the Age Discrimination in Employment Act (ADEA) is that the Title VII case is heard by a judge, while an ADEA case is heard by a jury).

One empirical study of the effect of statistical evidence on juries suggests that in general juries may tend to *underestimate* the value of statistical evidence, but when the probabilities described are quite small, this tendency decreases. David L. Faigman & A.J. Baglioni, Jr., *Bayes’ Theorem in the Trial Process: Instructing Jurors on the Value of Statistical Evidence*, 12 LAW & HUM. BEHAV. 1, 13–14 (1988).

272. See Kaye, *supra* note 33, at 1343–45, 1362 (arguing that because adoption of a particular level of statistical significance is arbitrary, testimony concerning whether results are significant or not should be supplanted by testimony concerning the size of the p-value).

of whether chance is a reasonable explanation for the observed disparity. Thus, although a p-value of .05 is an insufficient basis for concluding that the distribution probably did not occur by chance, a p-value of .00000001 may well support such an inference even if it is not an accurate measure of the probability that it was caused by nonchance factors, let alone caused by discrimination. However, the probative value in such a case is likely to be relatively slight; where disparities of that magnitude are present the primary focus of the litigation will probably be on identification of the systematic nonrandom cause rather than on whether the cause was random or nonrandom. Courts have repeatedly demonstrated the prejudicial effect of evidence of probabilities on themselves—allowing it inappropriately to skew their view of the case toward the plaintiff's theory—and that effect is likely to be amplified in jury trials. Because such evidence is so susceptible to misinterpretation, its prejudicial effect outweighs its limited probative value and it should not be admitted.

D. The Need for Proper Allocations of Burdens of Proof

Many courts have improperly allocated proof burdens in statistical cases, allowing a showing of little probative value to shift the burden of proof to the defendant. When, as frequently occurs, the evidence that the defendant is then called upon to produce is unavailable, this burden shift becomes outcome determinative. Such a result is unacceptable in a regime that places the ultimate burden on the plaintiff to demonstrate through competent evidence that the defendant has violated the law. As Brilmayer and Kornhauser have pointed out, "where opportunities for objective verification are scarce, adoption of a methodology takes on the characteristics of a value choice."²⁷³ Unfortunately, the value choices implicit in methods of statistical proof are inconsistent with the values of the statutes that they are designed to enforce.

Statistical evidence should no longer be allowed to serve as a basis for a finding of liability unless it can satisfy what are admittedly substantially more rigorous criteria than have heretofore been applied. This is not to require the "scientific certainty" abjured by the Supreme Court in *Bazemore*, but rather to require satisfaction of ordinary principles of evidence and proof burdens: a plaintiff must establish by a preponderance of reliable and relevant evidence that the employer has discriminated on a proscribed basis. Distracted by a perceived need to

273. Brilmayer & Kornhauser, *supra* note 39, at 152.

develop elaborate burden-shifting rules, courts have lost sight of this forest for the statistical trees.

If statistical proof of discrimination is still to be acceptable at all in court—which perhaps is doubtful—courts must pay more than lip service to the principle that throughout the litigation it is the plaintiff’s burden to demonstrate that impermissible discrimination is “the company’s standard operating procedure—the regular rather than the unusual practice.”²⁷⁴ The fundamental requirement must be that the plaintiff’s statistical analysis actually prove something. The plaintiff seeking to employ statistical analysis must demonstrate that all or substantially all of the factors that contribute to productivity—not just the minimum objective qualifications—are taken into account in the statistical model. Absent this demonstration, the statistical analysis, even if admissible under *Bazemore*, lacks sufficient probative value to establish a *prima facie* case.

Because employers do not hire at random from among the minimally qualified and because qualifications are not randomly distributed throughout the population, a showing that there is a statistical disparity after controlling for minimum qualifications has no probative value at all. Furthermore, because most employers do not base employment decisions purely on quantitative objective factors, and because there is no reason to believe that subjective qualifications vary throughout the population any less than objective factors, merely controlling for objective criteria is inadequate. Under a regime of logical rigor, plaintiffs would no longer be permitted to limit their focus to minimum qualifications for a job except in the unlikely event that they can demonstrate that all persons having the minimum qualifications would be equally qualified. Moreover, plaintiffs’ analyses could not be based upon unfounded (and often counterintuitive) assumptions of equality of interest and qualifications; the present judicial practice of accepting such assumptions and labelling defendants’ challenges to them as “speculative” would come to an end. In sum, the crude statistical showings that have to date formed the grist for the litigation mill should be recognized for what they are—capable of proving very little.

Admittedly, a plaintiff’s task would not be an easy one under a scheme imposing an appropriate measure of rigor; indeed, the burden would be insurmountable in many cases. That is not a reason, however, for declining to impose it.²⁷⁵ If shoddy statistics do not prove

274. *International Bhd. of Teamsters v. United States*, 431 U.S. 324, 336 (1977).

275. Brilmayer and Kornhauser, *supra* note 39, at 152, argue that allowing introduction of complex statistical analysis systematically favors large defendants, who have more ready access to experts. If an appropriate amount of rigor were applied to statistical evidence, there might be

discrimination by a preponderance of the evidence, so be it. There is no justification for lowering the standard simply because plaintiffs may otherwise not be able to meet it; that is a necessary consequence of placing the burden of proof on the plaintiff.²⁷⁶ One would have thought it obvious that if non-quantitative factors are important in the decision-making process—or if quantitative data are otherwise unavailable—it must follow that quantitative methods are an inadequate mode of analysis.²⁷⁷ Put another way—and we cannot ignore what may be a painful truth to some—many cases now treated as statistical cases are simply not amenable to statistical analysis.

V. CONCLUSION

Statistical proof of discrimination has been converted from what first appeared to be a common-sense application of a tool of inference into an elaborate game based upon ill-considered assumptions and presumptions.²⁷⁸ The primary method of analysis has been to construct a hypothetical abstract model of the employer's decision-making process and then to fit the data to the model. If the data and the model do not fit, it is assumed that the fault lies in the employment process rather than in the model. The fact that statistical analysis yields a number, carried out to as many decimal places as one pleases, has given courts an unjustified faith in the analysis. In the statistical sense, however, one must distinguish between precision and accuracy. The answer

a modicum of truth to the observation; however, the burden-shifting devices that have been described in this Article lessen to a large extent the need for plaintiffs to provide extensive technical presentations. Moreover, the recoverability of expert fees under the Civil Rights Act of 1991 increases the access of plaintiffs to such evidence. In any event, the plaintiff always may elect not to rely on statistical evidence, in which case disparities in access to statistical experts would not be an issue.

276. *But see* *Vuyanich v. Republic Nat'l Bank*, 505 F. Supp. 224, 358 (N.D. Tex. 1980) (rejecting the defendant's argument that general labor force comparisons are inappropriate for low-level clerical positions, reasoning, in part, that such a holding would "set up a standard of refinement at the *prima facie* stage . . . [in]consistent with the availability of statistical data"), *vacated*, 723 F.2d 1195 (5th Cir.), *cert. denied*, 469 U.S. 1073 (1984).

277. Judge Cudahy, dissenting in *EEOC v. Sears, Roebuck & Co.*, 839 F.2d 302, 365 (7th Cir. 1988), recognized that attempting to quantify the unquantifiable gives at best a crude approximation of reality. *Id.* (Cudahy, J., dissenting) ("To say that men are three times as interested as women in commission selling is a rough attempt to describe a dynamic social phenomenon; it is hardly a matter susceptible of precise delineation."). Nonetheless, he was willing to impose the burden of proof on the defendant to disprove the plaintiff's hypothetical number.

278. Richard Delgado has asserted that "computer-assisted analysis of data has enabled us to prove inequality more powerfully than ever before." Richard Delgado, *On Taking Back Our Civil Rights Promises: When Equality Doesn't Compute*, 1989 WIS. L. REV. 579, 579. As the foregoing discussion should reveal, the truth of Delgado's statement depends entirely on what he means by the word "prove."

provided by statistics may be very precise but it may give a wholly inaccurate sense of what is going on in the workplace. Answers, for answers' sake, are not the goal of the litigation process. After all, one can get an answer from a ouija board, and ouija boards are much cheaper than statistical experts.

Statistical models of an employer's decision-making process are valid only to the extent that they accurately capture the process.²⁷⁹ If they do not adequately reflect the process, they lack substantial probative value.²⁸⁰ Where courts have gone amiss in this area is that they often fail to consider defects in a plaintiff's statistical case to be the plaintiff's problem; instead, through a variety of devices they have made those defects the defendant's problem.

At bottom, the statistics relied upon by many courts reveal little about the probability that the defendant has engaged in a pattern or practice of discrimination, yet employers are routinely held liable on the basis of these meaningless statistics. The criticism of the many court opinions here should not be taken to mean that no courts are appropriately skeptical of statistical evidence. Some courts are,²⁸¹ but a proper amount of skepticism in cases where the proof is largely statistical would result, in virtually all cases, in a judgment for the defendant. Thus, the elaborate statistical presentations in pattern-or-

279. See *Sobel v. Yeshiva Univ.*, 839 F.2d 18, 22 (2d Cir. 1988) ("As with any multiple regression analysis, the validity of the influence attributed to a particular variable will depend heavily on how accurately the model mimics the actual factors influencing the dependent variable . . ."), *cert. denied*, 490 U.S. 1105 (1989); *Maddox v. Claytor*, 764 F.2d 1539, 1552 (11th Cir. 1985) ("If the tested disparity is based on erroneous assumptions or suffers from flaws in the underlying data, then standard deviation analysis is foredoomed to yield an equally faulty result."); *CONNOLLY ET AL.*, *supra* note 9, § 11.10 (pointing out the need for the statistical analysis to "reflect the process by which the employer's decisions were made"); *Brilmayer & Kornhauser*, *supra* note 39, at 120 ("[P]erfect technical implementation of a model does not guarantee scientific accuracy, since the most difficult part of quantitative modelling is not carrying out the computations, but selecting which computations to do"); *Campbell*, *supra* note 41, at 1305-12 (criticizing courts for their unquestioning acceptance of statistical models not shown to be a fair representation of the challenged employment process).

280. As the court stated in *Penk v. Oregon State Board of Higher Education*, 816 F.2d 458 (9th Cir.), *cert. denied*, 484 U.S. 853 (1987):

We note in passing that it is often acceptable in the social sciences to use a statistical model for proof of behavior even where, in an absolute sense, that model does not describe the data well. However, courts are not free to decide legal propositions on hypothetical evidence. This is especially true where courts are asked to draw inferences as to the existence of hidden discriminatory motives from statistical evidence.

Id. at 465 n.1. But see *Pitre v. Western Elec. Co.*, 843 F.2d 1262, 1269 (10th Cir. 1988) ("[S]tatistics that are insignificant to the social scientist may well be relevant to a court. . . . [W]hile social scientists search for certainty, the trier of fact in a Title VII case need only find that discrimination is more likely than not.")

281. See, e.g., *EEOC v. Chicago Miniature Lamp Works*, 947 F.2d 292 (7th Cir. 1991); *Sears, Roebuck & Co.*, 839 F.2d 302; *Coser v. Moore*, 739 F.2d 746 (2d Cir. 1984).

practice cases—such as the 20,000 pages of exhibits in the unsuccessful *Sears* case²⁸² or the “25,000 pages of testimony and rooms full of exhibits” in the unsuccessful *Penk* case²⁸³—are colossal wastes of time and resources. Moreover, the risk of liability based upon statistical evidence faced by even non-discriminating employers has the unfortunate consequence of pressuring employers to abandon legitimate, but unquantifiable, qualifications and to engage in hiring and promotion “by the numbers” to avoid liability under a statistical analysis. It is time to put an end to this misguided endeavor.

282. *Sears, Roebuck & Co.*, 839 F.2d at 307 n.2.

283. *Penk*, 816 F.2d at 460.